

Energy barriers and hysteresis in martensitic phase transformations

Zhiyong Zhang, Richard D. James and Stefan Müller

Abstract. We report results from a systematic program of changing composition of alloys in the system TiNiX , $X = \text{Cu, Pt, Pd, Au}$, to pursue certain special lattice parameters that have been identified previously with low hysteresis. We achieve $\lambda_2 = 1$, where λ_2 is the middle eigenvalue of the transformation strain matrix, for alloys with $X = \text{Pt, Pd, Au}$. In all cases there is a sharp drop of the graph of hysteresis vs. composition at the composition where $\lambda_2 = 1$. When the size of the hysteresis is replotted vs. λ_2 we obtain an universal graph for these alloys, which also agrees with trends in an earlier combinatorial study of alloys in the system TiNiCu . Motivated by these experimental results, we present a new theory for the size of the hysteresis based on the growth from a small scale of *fully developed* austenite martensite needles. In this theory the energy of the transition layer plays a critical role. New methods for calculation the optimal layer are developed that rely on Γ -convergence arguments, the small parameter being $|\lambda_2 - 1|$. The limiting energy of the transition layer is found to be governed by a nonstandard linear elasticity problem. Overall, the results point to a simple systematic method of achieving low hysteresis and a high degree of reversibility in transforming materials.

Contents

1	Introduction	2
2	Experimental methods	5
3	Theoretical methods	6
4	Measurements of hysteresis, transformation temperature and latent heat in alloys whose lattice parameters have been tuned to make $\lambda_2 = 1$	7
5	Background on the crystallographic theory and special relations among lattice parameters	11
6	Energy barrier of a fully developed austenite/martensite interface	13
7	Remarks and comparisons with other ideas about hysteresis	21

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE AUG 2008		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Energy barriers and hysteresis in martensitic phase transformations				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) California Institute of Technology 1200 E. California Blvd. Paasadena, CA				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 39	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

8	Theory of the transition layer between austenite and martensite when $\lambda_2 \sim 1$	23
8.1	Kinematic assumptions and boundary conditions	23
8.2	Change of variables	26
8.3	Scaling of the energy	27
8.4	Calculation of the Γ -limit of the energy	28
8.4.1	Hypotheses and summary of results used in the argument	28
8.4.2	Use of the rigidity lemma to restrict the form of a minimizer	29
8.4.3	Bounds on the positions f^\pm of the twin boundaries arising from energy minimization	31
8.4.4	Lower bound of the energy	32
8.4.5	Upper bound	34
8.5	Summary of form of the limiting energy of the transition layer	35

1 Introduction

This paper concerns the hysteresis that accompanies martensitic phase transformations. We focus mainly on thermal hysteresis.

The work reported here follows up a conjecture of [17] that asserts a relation between the conditions of compatibility between two phases and the hysteresis seen during cyclic transformation. The conjecture was formulated by looking at the literature (references in [17]) and trying to understand what is common among alloys with particularly low hysteresis. A special focus of this literature search was on data from repeated experiments on the same specimen, restored to its original shape after each test by heating, in a material with minor training effects, so as to factor out as much as possible processing conditions. Three conditions of compatibility were conjectured to be relevant to the minimization of hysteresis and the reversibility of transformation: 1) $\det \mathbf{U} = 1$, where the symmetric positive-definite matrix \mathbf{U} is the transformation strain matrix, 2) $\lambda_2 = 1$, where $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the ordered eigenvalues of \mathbf{U} , and 3) the conditions $\lambda_2 = 1$ and $\hat{\mathbf{a}} \cdot \mathbf{U}_1 \text{cof}(\mathbf{U}_1^2 - \mathbf{I})\mathbf{n} = 0$, together called the *cofactor conditions*. In the latter the vectors $\hat{\mathbf{a}}$ and \mathbf{n} are certain vectors that describe the twin system, as explained below (see (3)). The first of these conditions summarizes a well-accepted idea in the community.

Jun Cui et al. [8] noticed that the system $\text{Ti}_{50}\text{Ni}_{50-x}\text{Cu}_x$ for small values of x nearly satisfies the first two of these conditions, and they measured over a wide composition range both the eigenvalues of \mathbf{U} and the hysteresis, using combinatorial synthesis methods. Their results showed a strong correlation between the second of these conditions ($\lambda_2 = 1$) and the size of the hysteresis. Surprisingly, the measurements showed only a weak correlation between the size of the hysteresis and the volume change, even though $\det \mathbf{U}$ varied widely from about 0.96 to 1.07, with quite a few alloys near the extremes of this range. More surprisingly, the alloys with $\det \mathbf{U}$ very near 1 included some alloys with the largest, and others with the smallest hysteresis among all the alloys measured. All this is surprising because, since the earliest days of the study of martensite, the condition $\det \mathbf{U} \sim 1$ has been widely considered important for reversibility of the transformation, and elementary elasticity calculations of the stress field and energy surrounding an island of martensite growing in a hole of

austenite of a different volume suggests the presence of a rather large energy barrier associated with $\det \mathbf{U} \neq 1$.

The main purpose of this paper is two-fold: to describe the results of a program of alloy development in which the composition of alloys was systematically tuned to pursue the relation $\lambda_2 = 1$, and to give a new theory for the prediction of hysteresis.

We are able to achieve $\lambda_2 = 1$ in the systems TiNiPd, TiNiPt, and TiNiAu, and we exhibit alloys in each of these systems with $\lambda_2 < 1$ (like TiNi itself) and with $\lambda_2 > 1$. There is a sharp drop of the hysteresis for alloys with $\lambda_2 = 1$ with the most dramatic example being $\text{Ti}_{50}\text{Ni}_{37}\text{Au}_{13}$, which shows a decrease in the size of the hysteresis of about a factor of 10. When the size of the hysteresis is plotted as a function of composition, it has no particularly distinguishing features, but, when it is replotted as a function of λ_2 , there is a collapse of the data onto an approximately universal curve. The combinatorial data of Jun Cui et al. [8] also fall near this curve. This graph has some interesting features, including an apparent singularity at $\lambda_2 = 1$. This also is quite surprising, as hysteresis could well be expected to be sensitive to many other physical parameters besides λ_2 , as well as processing conditions.

To understand this dominance of λ_2 , we are led to propose a new theory of hysteresis. The condition $\lambda_2 = 1$ is the condition that there is an exact interface between austenite and martensite, and such an interface does not contribute the usual bulk energy of its transition layer, nor its interfacial energy on twin bands. The theory of hysteresis we propose is based on the hypothesis that the main energy barrier leading to hysteresis arises from these contributions. Thus, we propose that the main energy barrier leading to hysteresis is the growth of *fully developed* austenite/martensite interfaces. We implement this idea via a nucleation calculation. Ours is not the first nucleation calculation that has been done for martensite, but, as far as we can determine, previous calculations have not allowed such fully developed interfaces.

Recently, the interfaces between austenite and martensite in one of the alloys with $\lambda_2 = 1$ in the system TiNiPd has been observed using high resolution electron microscopy by Delville et al. [10]. These observations confirm the presence of untwinned, atomically sharp interfaces. The microstructures observed are unusual as compared with normal ($\lambda_2 \neq 1$) martensites.

This theory predicts a dramatic sensitivity of the size of the hysteresis to λ_2 . In particular, the predicted graph of hysteresis vs. λ_2 has a singularity at $\lambda_2 = 1$, like the measured data. The theory predicts that this graph depends on other physical parameters including the critical nucleus size, the other eigenvalues λ_1, λ_3 , the interfacial energy constant, an elastic modulus, the transformation temperature and the latent heat. However, for the alloys studied, the dependence on these other parameters, to the extent we can estimate them, is weak, consistent with the universality of the measured graphs. The graph of hysteresis vs. λ_2 is not predicted to be universal among all other alloys of other symmetries and properties.

Our prediction is not completely quantitative because of lack of knowledge of the interfacial energy constant, the critical nucleus size, and the fact that our model of the transition layer is not optimal. Because of recent and ongoing work [16, 32, 37, 38, 42, 45] on various new methods of measuring interfacial energy, progress is expected on the former.

Regarding the optimality of the transition layer, we present a new approach to the calculation of the optimal layer. We want to do this in order to lay the groundwork for a quantitative calculation of the hysteresis, but also for another reason suggested by this paper. That is, when we use a naive but reasonable calculation of the energy of the layer, we find a dramatic dependence of the energy of layer on the twin system that participates in the austenite/martensite interface, Table

4. For the lattice parameters of $\text{Ti}_{50}\text{Ni}_{50-x}\text{Pd}_x$, $x \sim 11$, this energy varies almost two orders of magnitude, depending on the twin system, even within a given twin type. This is quantified by a certain *geometric factor*. If this property extends to the optimal layer, then it could explain why certain twin systems are preferred in martensitic phase transformations.

Normally, the energy of the transition layer is thought to depend on the nonlinear elastic properties of the material. However, the elastic energy minimization problem for determining the structure of the layer contains a small parameter for these alloys, namely $|\lambda_2 - 1|$. This opens the way for the use of Γ -convergence arguments to derive a limiting variational principle for determination of the elastic energy of the layer for alloys with λ_2 near 1. We do this in Section 8. As is typical in such arguments the limiting variational principle has a universal status. That is, the input for the calculation of the limiting energy involves only the lattice parameters of austenite and martensite and the linear elastic moduli. In our derivation we consider only minimization of the elastic energy of the martensite. Future work should consider also relaxation of the austenite as well as the possibility of branching of the twins in the martensite [20, 21, 19]. We believe that our derivation lays the groundwork for these further calculations, which can lead to accurate evaluation of the energy in the transition layer. A side benefit of such a derivation is it will give a way to measure interfacial energy constant, by balancing bulk and interfacial energy in the usual way by studying alloys with λ_2 near 1 and using the measured fineness of the twins these alloys.

The full statement of the Γ -limiting problem is the problem of minimizing the following elastic energy,

$$\int_{\Omega_\ell} \frac{1}{2} \left[(\mathbf{v}_{,1} \otimes \mathbf{A}^{-T} \mathbf{m} + (\mathbf{a} + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}^{-T} \mathbf{n}) \cdot \mathbb{C} [\mathbf{v}_{,1} \otimes \mathbf{A}^{-T} \mathbf{m} + (\mathbf{a} + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}^{-T} \mathbf{n}] \right] \zeta dt_1 dt_2, \quad (1)$$

over functions \mathbf{v} and parameters $\hat{\eta}$ subject to the boundary conditions,

$$\begin{aligned} \mathbf{v}(t_1, 0) \parallel \mathbf{a}, \quad \mathbf{v}(t_1, 1) \parallel \mathbf{a}, \quad (\mathbf{v}(t_1, 0) - \mathbf{v}(t_1, 1)) \cdot \mathbf{a} \geq 0, \quad -\ell \leq t_1 \leq 0, \\ \mathbf{v}(0, t_2) = 0, \quad \mathbf{v}(-\ell, t_2) = \zeta(1 - t_2 + \hat{\eta})\mathbf{a}, \quad 0 \leq t_2 \leq 1. \end{aligned} \quad (2)$$

The notation is the following. Ω_ℓ is a rectangular domain of length ℓ , \mathbf{m} is the normal to the habit plane when the condition for exact compatibility ($\lambda_2 = 1$) is satisfied, \mathbf{a} is the amplitude of the twin and \mathbf{n} is its normal at $\lambda_2 = 1$, $\mathbf{A} = \mathbf{I} + \mathbf{b} \otimes \mathbf{m}$ is the deformation gradient of the martensite variant that is compatible with austenite at $\lambda_2 = 1$, \mathbb{C} is the elasticity tensor at \mathbf{A} , i.e., of the martensite at the compatible variant, ζ is the periodicity (length scale) of the twin bands, and $\hat{\eta}$ is a length that describes the offset of the vanishingly small twin band; essentially, it describes asymptotic bending of this band. It is believed that, without the rigorous Γ -convergence argument presented here, it would not be possible to guess this detailed form of this linear elasticity problem for the transition layer.

Reversibility, measured for example by fatigue life under cyclic transformation, is as important as low hysteresis for actuator applications. The connection between size of the initial or stabilized hysteresis loop and the fatigue life has been investigated by Gall and Maier [13], and Kato et al. [18], and recently by Moumni et al. [31]. All of these authors show a correlation between the hysteresis and fatigue life. The latter authors conclude that “it is shown that the dissipated energy of the stabilized cycle is a relevant parameter for the estimation of lifetime”.

There are many papers that discuss hysteresis in martensites. These include papers that propose constitutive equations for martensitic materials from which hysteresis loops are computed ([43],

[24], [6], [44], [26], [48]), those that discuss hysteresis in martensites in the presence of disorder ([39], [46]); papers that relate hysteresis to the attainment of a self-organized critical state [34], to metastability induced by incompatibility [2], and to pinning of interfaces by defects ([22], [49]). Another line of work models hysteresis using the Preisach model and its generalizations [4]. Not all of these theories are in direct contradiction with the one proposed here. Some of these theories are “micromechanical” and contain the parameter λ_2 that plays such an important role in this paper, and thus these theories can be implemented with λ_2 near 1 so as to examine their consistency with the measured hysteresis given here.

The notation is similar to that in the book of Bhattacharya [5]. In particular $\mathbf{a} \otimes \mathbf{n}$ is the 3×3 matrix with rectangular Cartesian components $a_i n_j$, defined from two vectors $\mathbf{a}, \mathbf{n} \in \mathbb{R}^3$. $\text{SO}(3) = \{\mathbf{R} \in \mathbb{R}^{3 \times 3} : \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det \mathbf{R} = 1\}$ denotes the set of 3×3 rotation matrices. Additional background for the Γ -convergence argument is given in Section 8.4.1.

2 Experimental methods

The main experimental results reported in this paper are measurements of hysteresis vs. lattice parameters in a family of TiNiX alloys, $X = \text{Cu, Pd, Pt, Au}$, in which the composition was systematically tuned to make the middle eigenvalue of the transformation strain matrix equal to 1. For the interpretation of these results in terms of energy barriers, measurements of transformation temperature, latent heat and density are needed, and these are also reported. The experimental results are presented in Section 4. In this section we briefly summarize the experimental methods that were used. More detailed information, particularly concerning uncertainty and the repeatability of the measurements, can be found in the thesis of Zhang [50].

Alloys of TiNiX, $X = \text{Cu, Pd, Pt, Au}$, were arc-melted on a water cooled copper hearth from high purity elemental materials under argon protection, after purging the chamber several times under vacuum. Ti was used as an hydrogen-getter. The resulting buttons were sliced using an electrical discharge machine (EDM) and then heat treated under vacuum in quartz ampoules, followed by a water quench. For the various compositions the annealing temperature was 700-850°C, and quenching water temperature was 0-40°C. A careful electrolytic polishing gave samples for both the x-ray measurements and differential scanning calorimetric (DSC) measurements. The electrolyte was 85-90% glacial acetic acid (CH_3COOH) and 10-15% perchloric acid (HClO_4) by volume, the cathode was stainless steel, the anode was stainless steel or Ti, the voltage was 35-40V, and the temperature of the bath was 0°C.

Lattice parameters of the alloys were measured on a Scintag x-ray diffractometer outfitted with a temperature controlled stage. Special attention was paid to alignment by using an internal standard (NIST standard reference 640c) and periodically doing in-situ alignment at different temperatures. The eigenvalues $\lambda_1, \lambda_2, \lambda_3$ were calculated directly from the lattice parameters and formulas given below. The density of the alloys was calculated from atomic composition of the starting materials and measured unit cell volume.

The DSC measurements were conducted on the TA Instruments Q1000 according to the ASTM standard F2004-03. The DSC samples were thinned to 100 μm , and electrolytically polished by the method described for the x-ray specimens.

Both the x-ray diffraction and calorimetry methods were used to determine the transformation temperatures. Austenite is stable at high temperature and martensite is stable at low temperature.

At some temperature θ_c in between, the two phases have the same bulk free energy. Because of the presence of hysteresis, θ_c is difficult to measure directly. When austenite is cooled down to a certain temperature, it begins to transform to martensite. This temperature is the martensite start temperature M_s . With further cooling, the transformation is completed at the martensite finish temperature M_f . When martensite is heated, it transforms back to austenite. The austenite start temperature and austenite finish temperature are denoted by A_s and A_f , respectively. To get these four temperatures from the DSC measurements we used the standard procedure of constructing intersections of approximate asymptotic lines. Operationally, we defined θ_c as the average of the four characteristic temperatures, i.e., $(A_s + A_f + M_s + M_f)/4$. In the x-ray measurements, the transformation temperature θ_c was defined as the intensity of austenite peak at half of its maximum value. The hysteresis was defined as $(A_s + A_f) - (M_s + M_f)$.

In the x-ray method we determined the temperatures M_s , M_f , A_s , A_f using the following procedure. The 2θ angle was first confined to the interval 39° to 46° because the (110) peak of B2 phase, the (111) peak of B19 phase, and the $(\bar{1}11)$, (111) peaks of the B19' phase of the TiNiX alloys, which are the strongest among all peaks of the corresponding phases, are within this range. We then scanned the sample at high temperature to obtain a pure austenite pattern. Using the same method, we scanned the sample at low temperature to get a pure martensite pattern. Then we gradually decreased the scan interval of temperature. Because x-ray patterns are sensitive to a change of lattice structures, they can detect a tiny phase change which would be difficult to quantify by other methods. To have a definite criterion for “start” and “finish” we established a convention. We let the pure austenite pattern at high temperature and the pure martensite pattern at low temperature be standards. Then, the temperatures at which the corresponding major peaks decreased 2% from those of the standard patterns were taken to be A_s or M_s , respectively. A_f and M_f were defined by a similar criterion using 98% of the standard peaks.

The latent heat L was defined as the area within the triangle with base $(A_f - A_s)/2$ and height equal to the maximum height of the DSC peak, divided by heating/cooling speed, which was typically 10°C/s . These were measured on the heating part of the cycle because the cooling part exhibited a longer tail that made the definition of M_f somewhat difficult.

3 Theoretical methods

The variable λ_2 will denote the middle eigenvalue of the transformation strain matrix. For example, in the case of a cubic to orthorhombic phase transformation, the six linear transformations that map the cubic structure to the six orthorhombic variants of martensite are given by (15) in the cubic basis. Their eigenvalues are assumed to be ordered, $\lambda_1 \leq \lambda_2 \leq \lambda_3$. The alloys of interest in this paper have λ_2 near 1.

The theoretical part of this paper relies on a technique called Γ -convergence, due in its abstract form to De Giorgi [9]. This method allows one to pass from a variational principle depending on a small parameter to a limiting variational principle. In the present case the small parameter is¹ $|\lambda_2 - 1|$, and the variational principle is the minimization of the free energy stored near the austenite/martensite interface. As is typical, the limiting variational principle is not obtained by putting $\lambda_2 = 1$ (or expanding the energy density in a Taylor series about $\lambda_2 = 1$). While one could

¹Or, equivalently, the small parameter can be chosen as λ , the volume fraction appearing in the crystallographic theory of martensite, this being related to $\lambda_2 - 1$ by (34).

imagine finding the minimizers of the original variational principle, calculating their asymptotic form as λ_2 approaches 1, and then seeking a variational principle for these asymptotic limits, the method of Γ -convergence gives the limit directly, without the intermediate step of calculating the minimizers of the original variational principle. The method involves the careful calculation of upper and lower bounds on low energy deformations as λ_2 tends to 1. It follows from the method that the limit, in a well-defined sense, of a sequence of minimizers (parameterized by λ_2) of the original variational principle is a minimizer of the limit energy, and their energies also converge to the minimum of the limiting energy.

As is typical with Γ -convergence, the limiting variational principle is simpler than the original one and has a universal status. In the present case this is realized by the fact that the limiting energy only depends on knowledge of the linearized elastic moduli of austenite and martensite, whereas the original principle depends on their nonlinear elastic properties. Once these moduli are measured for $\lambda_2 \sim 1$ alloys, and the interfacial energy on the twins are measured, the kinds of scaling laws discussed by Schryvers [38] can be investigated on a quantitative basis. Recent advances on the measurement of interfacial energies (see [37], [42], [45]) make this a realistic possibility.

4 Measurements of hysteresis, transformation temperature and latent heat in alloys whose lattice parameters have been tuned to make $\lambda_2 = 1$

For this study we began with the $\text{Ti}_{50}\text{Ni}_{50-x}\text{Cu}_x$ system, relying on a paper of Moberly and Melton [30] that showed the presence of twin-free habit planes in this alloy at $x = 10$. A closer examination of the original thesis of Moberly [29] revealed that these twin-free interfaces were rather rare: the photographs in the original thesis mainly showed twinned interfaces in the same alloy. In hindsight, we believe that the pictures chosen for publication were found in places in the specimen where, due to the presence of stress, the lattice parameters were perturbed from the stress-free, equilibrium values. (In our subsequent x-ray measurements this could be understood as arising from lattice parameters corresponding to points near the tails of the x-ray peaks in this system.) As is known, at about $x = 10$ at. % Cu this alloy loses the monoclinic martensite in favor of an orthorhombic martensite. We prepared a matrix of alloys of Ti = 48, 49, 50, 51, 52 at.%, Cu = 0, 5, 10, 15, 20, 25, 35 at.% with Ni = 100 - Ti - Cu and measured lattice parameters. The variation of λ_2 vs. composition for these alloys is shown in Figure 1.

For the case Ti = 50 at.% we also prepared alloys with Cu = 25, 30, 35 but these contained excessive TiCu precipitates and are not discussed here.

This data motivated the wider combinatorial study [8] of this system. Taken together, and accounting for the slightly decreased accuracy of the combinatorial measurements, we did not find convincing evidence that $\lambda_2 = 1$ was achieved at any composition in the TiNiCu system, although, as seen in Figure 1, $\text{Ti}_{49}\text{Ni}_{31}\text{Cu}_{20}$ comes quite close. Thus, we shifted further study to other alloys.

The Hume-Rothery rules for atomic size and valence indicated that Rh, Ir, Pd, Pt, Ag, and Au, all substituting for Ni, would be good candidates for stabilization of the orthorhombic phase. Extensive work on the alloys $\text{Ti}_{50}\text{Ni}_{50-x}\text{Ag}_x$ revealed a maximum solubility of $x = 2$ at. % Ag which we were not able to overcome by heat treatment, and therefore did not yield any useful alloys. We note that $\lambda_2 < 1$ is satisfied by all the TiNiCu alloys mentioned above, and also by

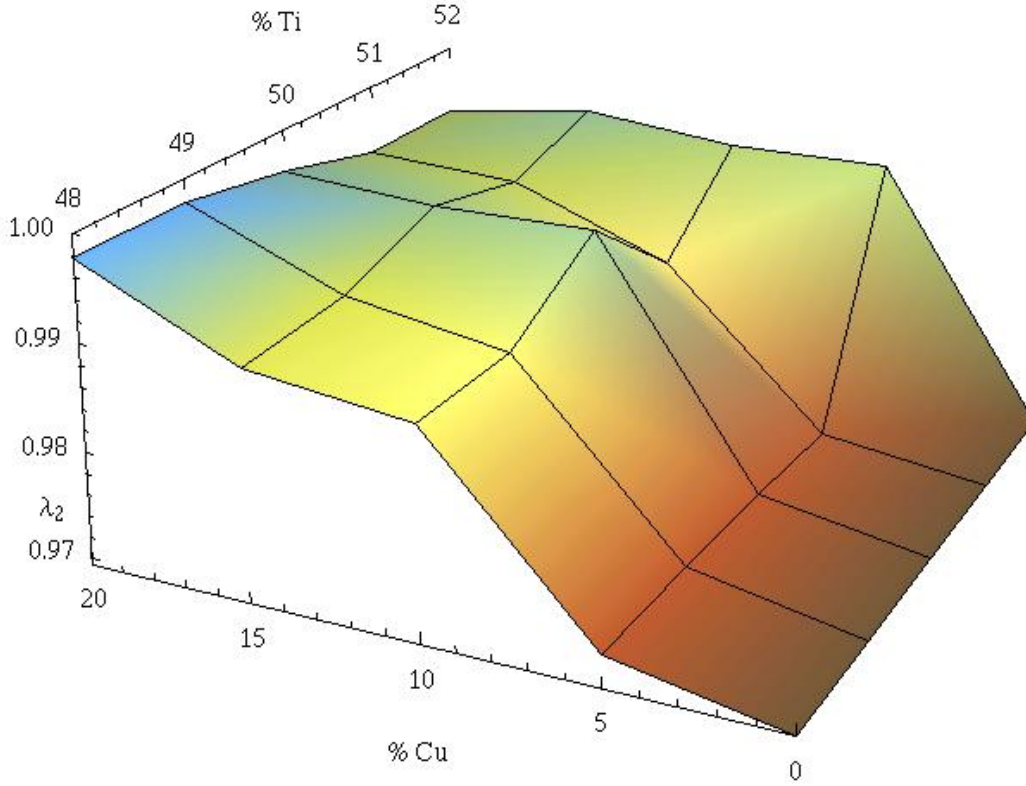


Figure 1: Measurements of λ_2 vs. composition in the TiNiCu system, Ti = 48, 49, 50, 51, 52 at.%, Cu = 0, 5, 10, 15, 20, 25, 35 at.% with Ni = 100 - Ti - Cu.

TiNi itself. Thus, the identification of starting alloys with $\lambda_2 > 1$ was critical for an interpolation scheme. We noticed from work of Donkersloot and van Vucht [11] that measured lattice parameters at the martensitic transformation in TiPd have $\lambda_2 > 1$, and our subsequent measurement of lattice parameters in $\text{Ti}_{50}\text{Ni}_{34}\text{Pd}_{16}$ showed $\lambda_2 > 1$. With regard to the basic hypothesis of a connection between hysteresis and $\lambda_2 = 1$, we were also motivated by the data collected in US Patent 5,951,793 which did not report lattice parameters but indicated a sharp drop in hysteresis for $\text{Ti}_{49.5}\text{Ni}_{40.5}\text{Pd}_{10}$. The review article of Miyazaki and Ishida [28] also indicated a sharp drop of the hysteresis near 10 at. % Pd in the data they report on films, although the interpretation suggested by the authors was of a continuous decrease of hysteresis with increasing Pd. Referring to their data, Miyazaki and Ishida state that this is a surprising result.

Having found that $\text{Ti}_{50}\text{Ni}_{34}\text{Pd}_{16}$ has $\lambda_2 = 1.005 > 1$ we did a straightforward interpolation. A similar procedure was followed with $\text{Ti}_{50}\text{Ni}_{50-x}\text{X}_x$, X = Pt, Au: first locate one alloy with $\lambda_2 > 1$, then interpolate backward. The data for hysteresis vs. composition for X = Pd, Pt, Au are shown in Figure 2. Note the drop of hysteresis of a factor between 5 and 10. Although the minimum hysteresis in all cases occurs at different values of x , in each case the minimum occurs very near the composition where $\lambda_2 = 1$.

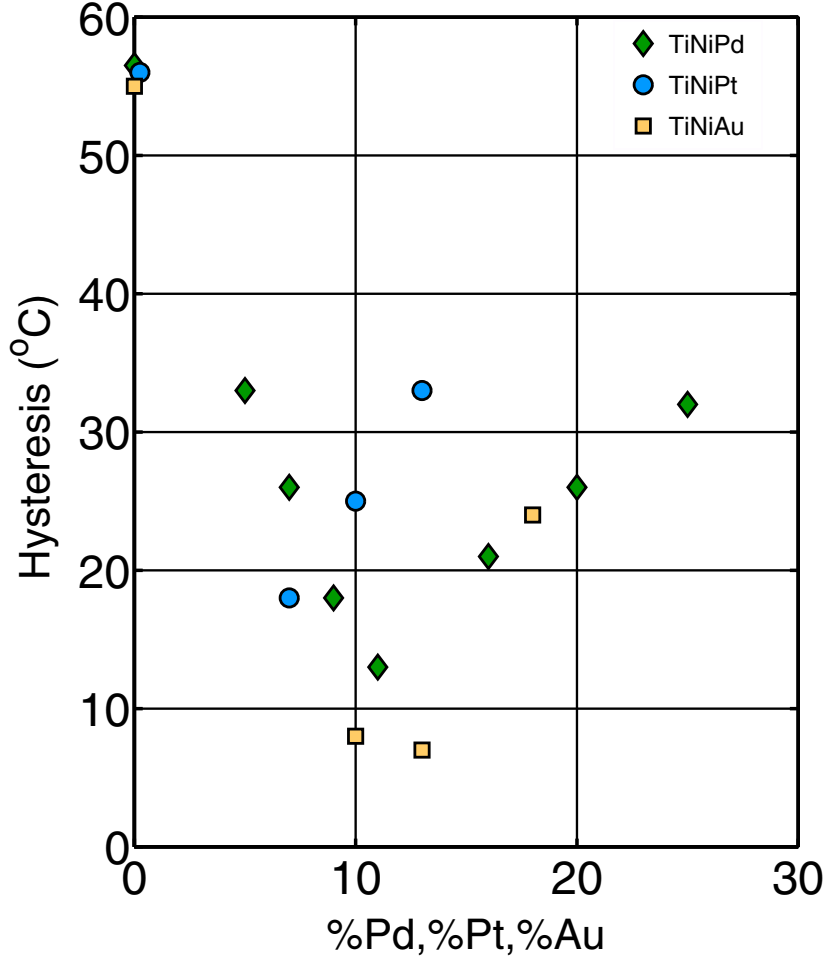


Figure 2: Width of the hysteresis vs. x for alloys $\text{Ti}_{50}\text{Ni}_{50-x}\text{X}_x$, $\text{X} = \text{Pd}, \text{Pt}, \text{Au}$.

At each data point of Figure 2 we also measured lattice parameters of the two phases at the same temperature. From this we calculated λ_2 corresponding to each point in Figure 2. It becomes more interesting to eliminate x and plot hysteresis directly vs. λ_2 . This is done in Figure 3. The combinatorial data from [8] on TiNiCu is also included in this plot. One can see a remarkable collapse of the data onto two lines shaped like a V. The suggestion of this plot is of a universal behavior. However, one should note that, while latent heats, elastic moduli and the other lattice parameters vary somewhat among these alloys, the crystallography is in all cases (except the TiNiHf data, see below) cubic to orthorhombic. Thus, one can conjecture that, within a certain crystallographic change, the main parameter that controls hysteresis is λ_2 and the behavior is universal. The theory presented later in this paper supports this viewpoint. We have also included on Figure 3 data on the system TiNiHf , which undergoes a cubic to monoclinic (B2 to B19') as in TiNi . The line is drawn using data of Matveeva et al. [27] and Potapov et al. [35] and was confirmed by our own measurements on this system. Alloys in the family TiNiHf , once considered promising candidates as high temperature shape memory alloys, show excessive hysteresis at concentrations of Hf that

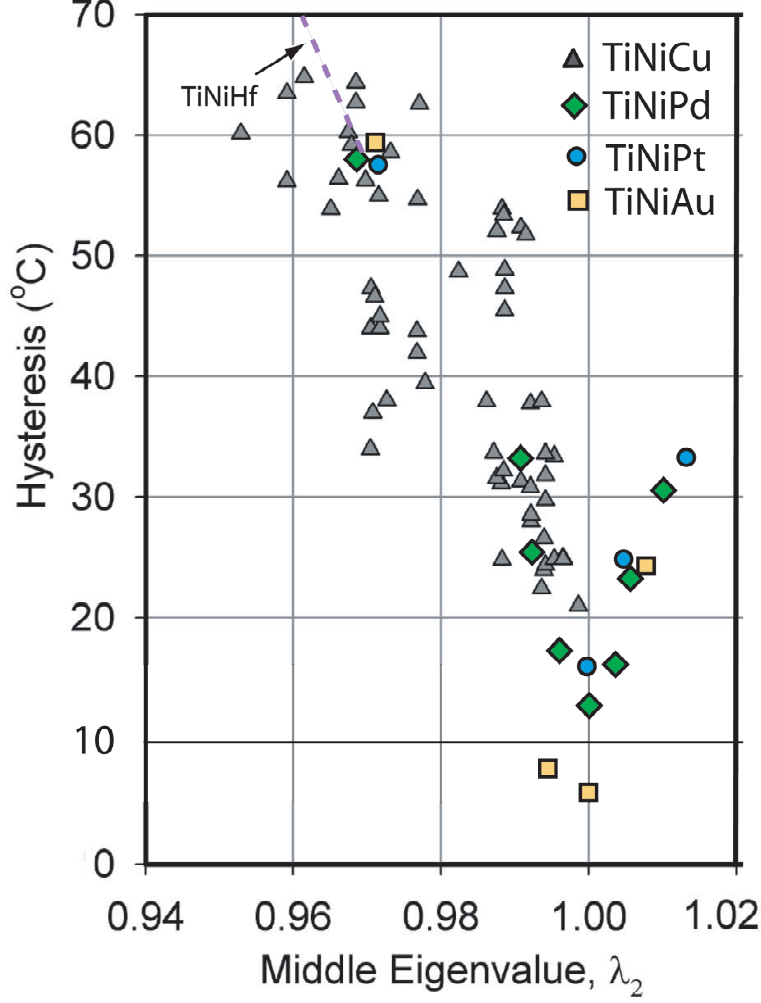


Figure 3: Width of the hysteresis vs. λ_2

raise the transformation temperature significantly. This can be understood from the following simple observation. Binary TiNi has $\lambda_2 < 1$. Increase of the percentage of Hf (substituted for Ti) *decreases* the value of λ_2 making it further from 1, significantly raising the hysteresis as indicated in Figure 3.

We note that we have previously reported [17] an extension of alloys satisfying $\lambda_2 = 1$ to the quaternary system TiNiCuPd.

For our later purpose of evaluating certain quantities in the theory, we will need the detailed data on lattice parameters (i.e., also λ_1 and λ_3) and latent heat. This data is reported in Table 1.

Table 1: Eigenvalues of the transformation strain matrix, latent heat and density of TiNiPd alloys

alloy	$\theta_c(^{\circ}\text{C})$	λ_1	λ_2	λ_3	latent heat (J/g)	density (g/cm ³)
TiNi	80				0.36	6.44
TiNi ₄₃ Pd ₇	-20	0.9398	0.9970	1.0606	0.30	6.75
TiNi ₄₁ Pd ₉	40	0.9341	0.9988	1.0635	0.14	6.76
TiNi ₃₉ Pd ₁₁	25	0.9280	1.0001	1.0674	0.23	6.85
TiNi ₃₄ Pd ₁₆	20	0.9280	1.0050	1.0633	0.38	7.11
TiNi ₃₂ Pd ₁₈	113	0.9227	1.0050	1.0710	0.58	7.20
TiNi ₂₈ Pd ₂₀	103	0.9244	1.0060	1.0691	0.56	7.35
TiNi ₂₈ Pd ₂₂	155	0.9186	1.0060	1.0753	0.70	7.40
TiNi ₂₅ Pd ₂₅	189	0.9167	1.0070	1.0775	0.60	7.54

5 Background on the crystallographic theory and special relations among lattice parameters

We consider a material transforming from austenite to martensite. Figure 4 sets the notation. We assume $|\mathbf{m}| = 1$, $|\mathbf{n}| = 1$ and of course \mathbf{m} not parallel to \mathbf{n} . The vectors \mathbf{n}^{\perp} and \mathbf{m}^{\perp} lie in the (\mathbf{m}, \mathbf{n}) plane and are uniquely determined by $\mathbf{n}^{\perp} \cdot \mathbf{n} = 0$, $\mathbf{m}^{\perp} \cdot \mathbf{m} = 0$, $|\mathbf{n}^{\perp}| = |\mathbf{m}^{\perp}| = 1$, $\mathbf{m}^{\perp} \cdot \mathbf{n} > 0$, $\mathbf{n}^{\perp} \cdot \mathbf{m} = \mathbf{m}^{\perp} \cdot \mathbf{n}$. (This normalization is not the standard one, but it is more convenient here.)

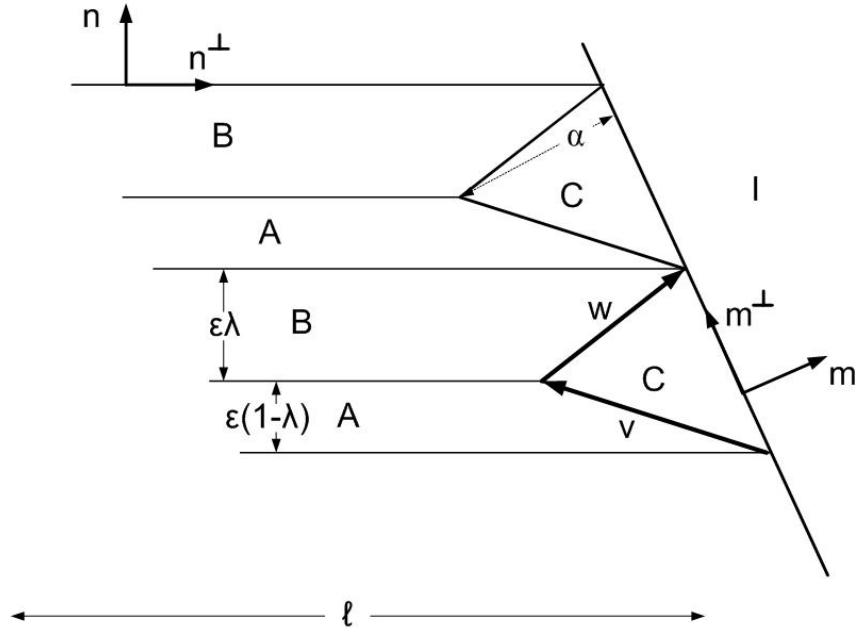


Figure 4: Austenite/martensite interface with a simple transition layer (C).

We assume a basic framework for martensitic phase transformations. That is, we consider a free energy density $\varphi(\mathbf{F}, \theta)$ as a function of deformation gradient and temperature defined on $M^{3 \times 3} \times \mathbb{R}^{\geq}$.

We hold θ fixed, below transformation temperature, and drop it from the notation until later in this paper. At this temperature φ is assumed to have a local minimum at \mathbf{I} and a lower global minimum on the martensite energy wells, $\text{SO}(3)\mathbf{U}_1, \dots, \text{SO}(3)\mathbf{U}_N$. We assume the distinct matrices $\mathbf{U}_1, \dots, \mathbf{U}_N$ are positive-definite, symmetric, and symmetry related: there exists $\tilde{\mathbf{R}}_2, \dots, \tilde{\mathbf{R}}_N \in \text{SO}(3)$ such that $\mathbf{U}_i = \tilde{\mathbf{R}}_i \mathbf{U}_1 \tilde{\mathbf{R}}_i^T, i = 2, \dots, N$. We also assume that the wells $\text{SO}(3)\mathbf{U}_1$ and $\text{SO}(3)\mathbf{U}_2$ are rank-1 connected, i.e., there are $\hat{\mathbf{a}}, \mathbf{n} \in \mathbb{R}^3, |\mathbf{n}| = 1$, and $\mathbf{R} \in \text{SO}(3)$ such that

$$\mathbf{R}\mathbf{U}_2 - \mathbf{U}_1 = \hat{\mathbf{a}} \otimes \mathbf{n}. \quad (3)$$

To do the asymptotic analysis we stay away from the singular situation $\lambda_2 = 1$ by assuming that $\det(\mathbf{U}_1^2 - \mathbf{I}) \neq 0$. The matrices $\mathbf{A} = \hat{\mathbf{R}}\mathbf{U}_1$ and $\mathbf{B} = \hat{\mathbf{R}}\mathbf{R}\mathbf{U}_2$, $\hat{\mathbf{R}} \in \text{SO}(3)$, are assumed to satisfy the equations of the crystallographic theory of martensite,

$$(\lambda\mathbf{B} + (1 - \lambda)\mathbf{A}) - \mathbf{I} = \mathbf{b} \otimes \mathbf{m}, \quad (4)$$

for suitable $0 < \lambda < 1$, $\mathbf{b}, \mathbf{m} \in \mathbb{R}^3, |\mathbf{m}| = 1$ and $\hat{\mathbf{R}} \in \text{SO}(3)$. Given the assumptions above, necessary and sufficient conditions that there is a solution $\hat{\mathbf{R}} \in \text{SO}(3)$, $\mathbf{b}, \mathbf{m} \in \mathbb{R}^3$ and $0 < \lambda < 1$ of (4) are that (see [3])

$$\begin{aligned} \delta \leq -2, \quad \text{and} \quad \text{tr}\mathbf{U}_1^2 - \det\mathbf{U}_1^2 + \frac{|\hat{\mathbf{a}}|^2|\mathbf{n}|^2}{2\delta} \geq 2, \\ \text{where } \delta = \hat{\mathbf{a}} \cdot \mathbf{U}_1(\mathbf{U}_1^2 - \mathbf{I})^{-1}\mathbf{n}. \end{aligned} \quad (5)$$

When these conditions are satisfied there are two solutions for λ given by $\lambda = \lambda^*$ and $\lambda = 1 - \lambda^*$ where:

$$\lambda^* = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{2}{\delta} + 1}. \quad (6)$$

Remark. (Brief summary of the crystallographic theory; see [47], [25], [7], and [3] for the version given here.) After substituting for \mathbf{B} using (3), the equation $(\lambda\mathbf{B} + (1 - \lambda)\mathbf{A}) - \mathbf{I} = \mathbf{b} \otimes \mathbf{m}$ becomes $\hat{\mathbf{R}}(\mathbf{U}_1 + \lambda\hat{\mathbf{a}} \otimes \mathbf{n}) = \mathbf{I} + \mathbf{b} \otimes \mathbf{m}$, which has a solution if and only if the middle eigenvalue of $\mathbf{G}_\lambda = (\mathbf{U}_1 + \lambda\mathbf{n} \otimes \hat{\mathbf{a}})(\mathbf{U}_1 + \lambda\hat{\mathbf{a}} \otimes \mathbf{n})$ is 1. This matrix has an eigenvalue equal to 1 if and only if

$$g(\lambda) = \det[(\mathbf{U}_1 + \lambda\mathbf{n} \otimes \hat{\mathbf{a}})(\mathbf{U}_1 + \lambda\hat{\mathbf{a}} \otimes \mathbf{n}) - \mathbf{I}] = 0. \quad (7)$$

Although it appears to be a sixth order polynomial, $g(\lambda)$ is actually quadratic in λ and symmetric about $1/2$, so it can be written

$$g(\lambda) = \zeta(\lambda - \frac{1}{2})^2 + \eta. \quad (8)$$

The simplest evaluation of ζ and η follows from the equations

$$\begin{aligned} \frac{1}{4}\zeta + \eta &= \det(\mathbf{G}_\lambda - \mathbf{I})|_{\lambda=0} = \det(\mathbf{U}_1^2 - \mathbf{I}) \\ -\zeta &= \frac{d}{d\lambda} \det(\mathbf{G}_\lambda - \mathbf{I})|_{\lambda=0} = 2(\det(\mathbf{U}_1^2 - \mathbf{I}))(\hat{\mathbf{a}} \cdot \mathbf{U}_1(\mathbf{U}_1^2 - \mathbf{I})^{-1}\mathbf{n}) \end{aligned} \quad (9)$$

The condition that $g(\lambda)$ has a root on $(0, 1)$ is that $g(0)g(1/2) \leq 0$ which gives the first of (5). Suppose this holds and call the roots $\lambda^*, 1 - \lambda^*$. A necessary and sufficient condition that these roots are associated to the middle eigenvalue is that

$$0 \leq (\lambda_1 - 1)(1 - \lambda_3) = \lambda_1 + \lambda_3 - 1 - \lambda_1\lambda_3 = \text{tr}\mathbf{G}_{\lambda^*} - 2 - \det\mathbf{G}_{\lambda^*}, \quad (10)$$

which gives the second of (5). We note that the *cofactor conditions* discussed in [17] are the conditions that i) the middle eigenvalue of \mathbf{U}_1 is 1 and ii) $g(\lambda)$ is identically 0, i.e., that both ζ and η vanish. In this case $\det(\mathbf{U}_1^2 - \mathbf{I}) = 0$, and the second equation in (9) remains valid in the modified form $\zeta = -2\hat{\mathbf{a}} \cdot \mathbf{U}_1 \text{cof}(\mathbf{U}_1^2 - \mathbf{I})\mathbf{n} = 0$.

We let $\mathbf{a} = \hat{\mathbf{R}}\hat{\mathbf{a}}$ with $\hat{\mathbf{R}}$ obtained as in this remark, that is, $\mathbf{B} - \mathbf{A} = \mathbf{a} \otimes \mathbf{n}$. A simple transition layer is the one shown in Figure 4. It is kinematically possible because, using (5), there is a matrix \mathbf{C} such that

$$\mathbf{C} - \mathbf{I} = \mathbf{f} \otimes \mathbf{m}, \quad \mathbf{C}\mathbf{v} = \mathbf{A}\mathbf{v}, \quad \mathbf{C}\mathbf{w} = \mathbf{B}\mathbf{w}. \quad (11)$$

(see Figure 4). To verify this, note first that

$$\mathbf{v} = \frac{1}{\mathbf{n} \cdot \mathbf{m}^\perp} ((1 - \lambda)\varepsilon\mathbf{m}^\perp - \alpha\mathbf{n}^\perp), \quad \mathbf{w} = \frac{1}{\mathbf{n} \cdot \mathbf{m}^\perp} (\lambda\varepsilon\mathbf{m}^\perp + \alpha\mathbf{n}^\perp) \quad (12)$$

for some $\alpha > 0$, which is the height of the triangle as shown in Figure 4 as one can see by computing $-\mathbf{v} \cdot \mathbf{m}$.

Now it can be seen that the following expression for \mathbf{C} satisfies (11):

$$\mathbf{C} = \mathbf{I} + \mathbf{f} \otimes \mathbf{m}, \quad \mathbf{f} = \mathbf{b} + \frac{\varepsilon}{\alpha}\lambda(1 - \lambda)\mathbf{a}. \quad (13)$$

The matrix \mathbf{C} is of course not on the energy wells.

6 Energy barrier of a fully developed austenite/martensite interface

Consider the case of cooling from austenite to form martensite. We picture a lenticular nucleus, stabilized by a defect (e.g., precipitate, triple junction on a grain boundary), and we imagine that this nucleus is already present at temperatures above M_s . Our main hypothesis is that this nucleus is already twinned, so its boundary consists of two fully developed, slightly bowed austenite/martensite interfaces. We ignore the ends of the nucleus, and picture simply a twinned laminate of martensite of length ℓ (i.e., the half-width of the nucleus) meeting austenite. As sources of energy, we consider the interfacial energy on the twin boundaries modeled by a sharp interface theory, the bulk elastic energy stored near the austenite/martensite interface, and the bulk free energies of the two phases. We analyze the energy as a function of ℓ accounting for the possibility that the twin density and elastic transition layer may relax as ℓ is changed. In this section we consider a simple model for the transition layer that is improved later in this paper.

We consider the case in which the middle eigenvalue λ_2 of the matrix \mathbf{U}_1 is close to 1 and, for simplicity, we fix the other eigenvalues². In a basis for which \mathbf{U}_1 is diagonal, we then evaluate the formula (5) for δ :

$$\delta = (a_1 n_1) \frac{\lambda_1}{\lambda_1^2 - 1} + (a_2 n_2) \frac{\lambda_2}{\lambda_2^2 - 1} + (a_3 n_3) \frac{\lambda_3}{\lambda_3^2 - 1}. \quad (14)$$

²The other eigenvalues λ_1, λ_3 could also be allowed to vary as long as they are bounded away from 1; this of course forbids the cubic to tetragonal case, since in that case $\lambda_2 = 1$ and $\lambda_1 \leq \lambda_2 \leq \lambda_3$ imply that $\lambda_1 = 1$ or $\lambda_3 = 1$.

This is dominated by the middle term and we note that, typically, as λ_2 approaches 1, the quantity a_2n_2 does not tend to zero³. In view of the first condition of (5) we have the following: **assuming that $a_2n_2|_{\lambda_2=1} \neq 0$ then for a given twin system the crystallographic theory has solutions for at most one of the regions $\lambda_2 \gtrsim 1$ or $\lambda_2 \lesssim 1$.** (See also Items 1 and 2 below.)

We now give a more detailed study of this in the cubic-to-orthorhombic case, which applies to all the NiTiX alloys described above. In this case the six transformation strain matrices are,

$$\begin{aligned} \mathbf{U}_1 &= \begin{pmatrix} \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_2-\lambda_3}{2} & 0 \\ \frac{\lambda_2-\lambda_3}{2} & \frac{\lambda_2+\lambda_3}{2} & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}, & \mathbf{U}_2 &= \begin{pmatrix} \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_3-\lambda_2}{2} & 0 \\ \frac{\lambda_3-\lambda_2}{2} & \frac{\lambda_2+\lambda_3}{2} & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}, \\ \mathbf{U}_3 &= \begin{pmatrix} \frac{\lambda_2+\lambda_3}{2} & 0 & \frac{\lambda_2-\lambda_3}{2} \\ 0 & \lambda_1 & 0 \\ \frac{\lambda_2-\lambda_3}{2} & 0 & \frac{\lambda_2+\lambda_3}{2} \end{pmatrix}, & \mathbf{U}_4 &= \begin{pmatrix} \frac{\lambda_2+\lambda_3}{2} & 0 & \frac{\lambda_3-\lambda_2}{2} \\ 0 & \lambda_1 & 0 \\ \frac{\lambda_3-\lambda_2}{2} & 0 & \frac{\lambda_2+\lambda_3}{2} \end{pmatrix}, \\ \mathbf{U}_5 &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_2-\lambda_3}{2} \\ 0 & \frac{\lambda_2-\lambda_3}{2} & \frac{\lambda_2+\lambda_3}{2} \end{pmatrix}, & \mathbf{U}_6 &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_2-\lambda_3}{2} \\ 0 & \frac{\lambda_2-\lambda_3}{2} & \frac{\lambda_2+\lambda_3}{2} \end{pmatrix}. \end{aligned} \quad (15)$$

Written this way $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of each of these matrices, and, for the application to the TiNiX alloys studied here, we can without loss of generality assume these are ordered, $\lambda_1 \leq \lambda_2 \leq \lambda_3$. We consider twin systems that consist of variant 1 twinned with all other variants; it is sufficient to consider these by symmetry. Variant 1 is compound twinned with variant 2, but it forms Type I/Type II twins with the other four variants. (Thus, all pairs of variants are twinned). By direct calculation, in the notation of (14), we have the expressions for a_2n_2 listed in Table 2 at $\lambda_2 = 1$.

Table 2: Values of a_2n_2 for various twin systems at $\lambda_2 = 1$. Cubic-to-orthorhombic case.

	twins between variants 1 and 2	twins between variants 1 and 3, 4, 5, 6
value of a_2n_2 (Type I twins)	$\frac{\lambda_3^2-1}{\lambda_3^2+1}$	$\frac{\lambda_1^2(3\lambda_3^2-1)-2\lambda_3^2}{2(2\lambda_3^2+\lambda_1^2(1+\lambda_3^2))}$
value of a_2n_2 (Type II twins)	$\frac{\lambda_3^2-1}{\lambda_3^2+1}$	$\frac{\lambda_3^2+2\lambda_1^2-3}{2(\lambda_3^2+2\lambda_1^2+1)}$

The numerators of the expressions in the last column of Table 2, for eigenvalues near 1, are close to $\lambda_3^2 - 1 + 2(\lambda_1^2 - 1)$ which is typically negative for experimental cases and also for any case which is nearly volume preserving. Hence, under this typical situation that entries in the last column of Table 2 are negative, we reach the following conclusions:

1. $\lambda_2 \lesssim 1$. Only compound twins form austenite/martensite interfaces. Altogether, there are only six of these twin systems, two each for variants 1-2, 3-4, 5-6.
2. $\lambda_2 \gtrsim 1$. All of the Type I and Type II twin systems can form austenite/martensite interfaces.

³“Typically”, here means that if \mathbf{U}_1 is changed to $\rho\mathbf{U}_1$ for ρ near 1, the middle eigenvalue of $\rho\mathbf{U}_1$ can be made to pass through 1 but a_2n_2 is only changed to ρa_2n_2 .

This striking conclusion is consistent with results of Hane and Shield [14]. Experimental values of the dominant factor $a_2 n_2$ for alloys in which $\lambda_2 = 1$ are given in Table 3. Within a given twin type, these numbers do not vary greatly among these various systems.

Table 3: Values of $a_2 n_2$ for alloys of TiNiX, X = Cu, Pt, Pd, Au, at compositions where $\lambda_2 = 1$

	twins between var. 1 and 2	twins between var. 1 and 3, 4, 5, 6
Ti ₅₀ Ni ₂₆ Cu ₂₀ Pd ₄ (Type I twins)	0.0534	-0.0137
Ti ₅₀ Ni ₂₆ Cu ₂₀ Pd ₄ (Type II twins)	0.0534	-0.0103
Ti ₅₀ Ni ₃₉ Pd ₁₁ (Type I twins)	0.0651	-0.024
Ti ₅₀ Ni ₃₉ Pd ₁₁ (Type II twins)	0.0651	-0.018
Ti ₅₀ Ni ₄₃ Pt ₇ (Type I twins)	0.0645	-0.0212
Ti ₅₀ Ni ₄₃ Pt ₇ (Type II twins)	0.0645	-0.0157
Ti ₅₀ Ni ₃₇ Au ₁₃ (Type I twins)	0.0606	-0.018
Ti ₅₀ Ni ₃₇ Au ₁₃ (Type II twins)	0.0606	-0.0133

One should remark that even though the crystallographic theory forbids certain twins to participate in the austenite/martensite interface, all the types of twins listed above are possible in the martensite. For example, in cases where compound twins are forbidden by the crystallographic theory to be present in austenite/martensite interfaces, they could still be observed in the martensite, produced by stresses built up during transformation, especially in a polycrystal.

Assume that we are in one of the regions where the crystallographic theory has a solution, that is, the second of (5) holds and either $\lambda_2 \gtrsim 1$ and $a_2 n_2 < 0$ or $\lambda_2 \lesssim 1$ and $a_2 n_2 > 0$. From (6) we then have that in the regime of interest in this paper $\lambda^* \sim 1$. We confine attention to the solution $\lambda = 1 - \lambda^*$. Thus, the volume fraction $\lambda \sim 0$, i.e., it is the volume fraction of the variant that is paired with variant 1 that is disappearing as $\lambda_2 \rightarrow 1$.

Thus the martensite region contains a large volume fraction where the deformation gradient is \mathbf{A} . Also, from equations (4) and (13) we have

$$\mathbf{A} \rightarrow \mathbf{I} + \mathbf{b} \otimes \mathbf{m}, \quad \mathbf{C} \rightarrow \mathbf{I} + \mathbf{b} \otimes \mathbf{m} \quad (16)$$

as $\lambda_2 \rightarrow 1$ with ε/α fixed. Since \mathbf{C} is near \mathbf{A} in this limit, it is natural to estimate the energy of \mathbf{C} using a quadratic approximation of the free energy density φ near \mathbf{A} . We would also like to optimize over the choice of α (see Fig. 4), but we keep ε fixed for now. Later, after we have introduced interfacial energy, we will optimize over the choice of the “fineness” ε .

Note that $\mathbf{C} = \mathbf{I} + \mathbf{f} \otimes \mathbf{m} = \mathbf{I} + \mathbf{b} \otimes \mathbf{m} + (\varepsilon/\alpha)\lambda(1 - \lambda)\mathbf{a} \otimes \mathbf{m}$, and from (4) we have $\mathbf{A} =$

$\mathbf{I} + \mathbf{b} \otimes \mathbf{m} - \lambda \mathbf{a} \otimes \mathbf{n}$. Thus, as must be true, $\text{rank}(\mathbf{C} - \mathbf{A}) = 1$, i.e.,

$$\mathbf{C} = \mathbf{A} + \mathbf{a} \otimes \left[\lambda \mathbf{n} + \frac{\varepsilon}{\alpha} \lambda (1 - \lambda) \mathbf{m} \right]. \quad (17)$$

For our simple estimate of the energy of \mathbf{C} we let μ be a typical elastic shear modulus, we note that the area of each triangle containing \mathbf{C} is

$$\frac{\varepsilon \alpha}{2 \mathbf{m} \cdot \mathbf{n}^\perp}. \quad (18)$$

(Recall that by definition $\mathbf{m} \cdot \mathbf{n}^\perp > 0$.) At this point it is convenient to fix one value of the free energy by putting $\varphi(\mathbf{A}, \theta) = 0$, so the free energy of \mathbf{C} will be positive. This of course implies that $\varphi(\mathbf{I}, \theta) > 0$. Suppose the width (into the page) of the martensite plate is w . Then, we can estimate the energy of \mathbf{C} by using a linear elasticity theory⁴ obtained by linearizing about \mathbf{A} . Using (18) and letting $\mathbf{c}_\lambda = \mathbf{A}^{-T} \mathbf{n} + (\varepsilon/\alpha)(1 - \lambda) \mathbf{A}^{-T} \mathbf{m}$ (so that $\mathbf{C} \mathbf{A}^{-1} = \mathbf{I} + \lambda \mathbf{a} \otimes \mathbf{c}_\lambda$), we have

$$\begin{aligned} \text{energy of one triangle} &= \frac{\varepsilon \alpha w}{2 \mathbf{m} \cdot \mathbf{n}^\perp} \left(\frac{\mu}{2} \left| \frac{1}{2} [(\mathbf{C} \mathbf{A}^{-1} - \mathbf{I}) + (\mathbf{C} \mathbf{A}^{-1} - \mathbf{I})^T] \right|^2 \right) \\ &= \frac{\varepsilon \alpha w}{2 \mathbf{m} \cdot \mathbf{n}^\perp} \left(\frac{\mu \lambda^2}{8} |\mathbf{a} \otimes \mathbf{c}_\lambda + \mathbf{c}_\lambda \otimes \mathbf{a}|^2 \right) \\ &= \frac{\varepsilon \alpha w}{2 \mathbf{m} \cdot \mathbf{n}^\perp} \left(\frac{\mu \lambda^2}{4} ((\mathbf{a} \cdot \mathbf{c}_\lambda)^2 + |\mathbf{a}|^2 |\mathbf{c}_\lambda|^2) \right). \end{aligned} \quad (19)$$

The triangle height α is contained in the prefactor as well as in the expression for \mathbf{c}_λ . Minimization of this energy over $\alpha > 0$ gives

$$\alpha_{\min} = \varepsilon(1 - \lambda) \sqrt{\frac{(\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{m})^2 + |\mathbf{a}|^2 |\mathbf{A}^{-T} \mathbf{m}|^2}{(\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{n})^2 + |\mathbf{a}|^2 |\mathbf{A}^{-T} \mathbf{n}|^2}}. \quad (20)$$

We insert this value of α from now on. Then, assuming the height of the martensite plate is h , giving h/ε triangles, we then have

$$\text{total energy of the triangles} = \varepsilon w h \mu \lambda^2 (1 - \lambda) \xi, \quad (21)$$

where the *geometric factor* ξ is given by

$$\begin{aligned} \xi &= \frac{1}{4 \mathbf{m} \cdot \mathbf{n}^\perp} \left(\sqrt{\left((\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{m})^2 + |\mathbf{a}|^2 |\mathbf{A}^{-T} \mathbf{m}|^2 \right) \left((\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{n})^2 + |\mathbf{a}|^2 |\mathbf{A}^{-T} \mathbf{n}|^2 \right)} \right. \\ &\quad \left. + (\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{m})(\mathbf{a} \cdot \mathbf{A}^{-T} \mathbf{n}) + |\mathbf{a}|^2 (\mathbf{A}^{-T} \mathbf{m} \cdot \mathbf{A}^{-T} \mathbf{n}) \right) \end{aligned} \quad (22)$$

The geometric factor measures in a rough way, i.e., for a simple model of the transition layer, the effect that different twin systems have on the elastic energy stored in the austenite/martensite

⁴Given that the crystallographic theory is geometrically exact, one could worry about the possible role of finite rotations in this calculation, and therefore be led to use nonlinear elasticity instead of linear elasticity to estimate the energy of \mathbf{C} . The Γ -convergence argument given later in this paper justifies the use of linearized elasticity (linearized about \mathbf{A}) in this context for λ near 1.

interface. It depends only on the data that goes into the crystallographic theory, that is, the energy wells and lattice parameters. It is interesting to look at values of the geometric factor for various twin systems. This is done in Table 4 for the alloy $\text{Ti}_{50}\text{Ni}_{50-x}\text{Pd}_x$, $x \sim 11$. It can be seen that there is a very wide variation of almost two orders of magnitude of this number, and it also varies significantly within a given type of twin. This partly motivated the more careful treatment of the transition layer given later in this paper. It may possibly explain why certain twins predominate in certain alloys. It also suggests that within a certain twin type, certain pairs of variants would tend to be rather strongly preferred.

Table 4: Geometric factor ξ for various twin systems for $\text{Ti}_{50}\text{Ni}_{50-x}\text{Pd}_x$, $x \sim 11$. Note the wide variation. Components in the cubic basis.

variants	twin type	habit plane	$\xi \times 10^3$
1 and 2	compound	(-0.5005, 0.5005, -0.7064)	8.07 3.47
1 and 3	Type I Type II	(-0.5005, 0.5005, -0.7064)	48.6 19.8
1 and 4	Type I Type II	(-0.5005, 0.5005, -0.7064)	14.7 2.07
1 and 5	Type I Type II	(-0.5005, 0.5005, -0.7064)	23.4 162
1 and 6	Type I Type II	(-0.5005, 0.5005, -0.7064)	3.64 16.9
1 and 2	compound	(0.5005, -0.5005, -0.7064)	3.47 8.07
1 and 3	Type I Type II	(0.5005, -0.5005, -0.7064)	23.4 162
1 and 4	Type I Type II	(0.5005, -0.5005, -0.7064)	3.64 16.9
1 and 5	Type I Type II	(0.5005, -0.5005, -0.7064)	48.6 19.8
1 and 6	Type I Type II	(0.5005, -0.5005, -0.7064)	14.7 2.07

Now we repeat the classical calculation with this model of the transition layer, except we have the austenite and martensite wells at different heights. We introduce an interfacial energy per unit area κ on the twins and write the total energy as

$$\text{total energy} = \frac{2\kappa wh\ell}{\varepsilon} + \varepsilon wh\mu\lambda^2(1 - \lambda)\xi + \varphi(\mathbf{A}, \theta)wh\ell + \varphi(\mathbf{I}, \theta)wh(L - \ell), \quad (23)$$

where L is the total length of the strip, including martensite and austenite ($L \gg \ell$). We minimize

this expression over $\varepsilon > 0$ and get

$$\varepsilon = \sqrt{\frac{2\kappa\ell}{\mu\lambda^2(1-\lambda)\xi}}, \quad (24)$$

giving an inverse relation between twin spacing ε and volume fraction λ in the limit as $\lambda \rightarrow 0$ (i.e., as $\lambda_2 \rightarrow 1$). This gives the total minimum energy

$$2wh\lambda\sqrt{2\kappa\mu\ell(1-\lambda)\xi} + wh\ell(\varphi(\mathbf{A}, \theta) - \varphi(\mathbf{I}, \theta)) + \text{const.} \quad (25)$$

This has the typical form of an energy with a single barrier that arises in a nucleation calculation. The $\sqrt{\ell}$ term dominates for small ℓ and the linear term dominates for large ℓ . Recall that we have assumed that $0 = \varphi(\mathbf{A}, \theta) < \varphi(\mathbf{I}, \theta)$; thus, the energy goes to $-\infty$ as $\ell \rightarrow \infty$. Focusing now on the dependence of (25) on ℓ , we see that the energy (25) has a single maximum at

$$\ell_{\max} = \frac{2\lambda^2\kappa\mu(1-\lambda)\xi}{(\varphi(\mathbf{A}, \theta) - \varphi(\mathbf{I}, \theta))^2}. \quad (26)$$

The value of the total energy at ℓ_{\max} is

$$\text{energy at the barrier} = \frac{2\lambda^2wh\kappa\mu(1-\lambda)\xi}{\varphi(\mathbf{I}, \theta) - \varphi(\mathbf{A}, \theta)}. \quad (27)$$

Note that the energy at the barrier and ℓ_{\max} are both proportional to λ^2 . Thus, as $\lambda_2 \rightarrow 1$ both ℓ_{\max} and the height of the barrier tend to 0 at the same rate.

Now we can make a simple criterion based on the idea that typical defects in the material stabilize fully developed austenite/martensite interfaces at a value $\ell = \ell_c$. Then such interfaces are expected to move when $\varphi(\mathbf{I}, \theta) - \varphi(\mathbf{A}, \theta)$ is increased (by decreasing temperature) to a point where ℓ_{\max} given by (26) exceeds ℓ_c . This gives the criterion for the size of the hysteresis,

$$\frac{2\lambda^2\kappa\mu(1-\lambda)\xi}{(\varphi(\mathbf{A}, \theta) - \varphi(\mathbf{I}, \theta))^2} = \ell_c. \quad (28)$$

To evaluate the denominator of this expression, consider a Taylor expansion of each energy well about $\theta = \theta_c$. The minimizers \mathbf{A}, \mathbf{I} on the martensite and austenite wells, respectively, depend weakly on temperature, so we expand the notation by writing $\mathbf{A}(\theta), \mathbf{I}(\theta)$; these are still chosen to lie on the wells and to satisfy the equations of the crystallographic theory as given above. Expanding φ at each well about $\theta = \theta_c$ we have

$$\begin{aligned} \varphi(\mathbf{A}(\theta), \theta) &= \varphi(\mathbf{A}(\theta_c), \theta_c) + \left(\frac{\partial\varphi(\mathbf{A}(\theta^*), \theta^*)}{\partial\theta} + \frac{\partial\varphi(\mathbf{A}(\theta^*), \theta^*)}{\partial\mathbf{F}} \cdot \mathbf{A}'(\theta^*) \right) (\theta - \theta_c), \\ \varphi(\mathbf{I}(\theta), \theta) &= \varphi(\mathbf{I}(\theta_c), \theta_c) + \left(\frac{\partial\varphi(\mathbf{I}(\theta^*), \theta^*)}{\partial\theta} + \frac{\partial\varphi(\mathbf{I}(\theta^*), \theta^*)}{\partial\mathbf{F}} \cdot \mathbf{I}'(\theta^*) \right) (\theta - \theta_c), \end{aligned} \quad (29)$$

where θ^* is a suitable intermediate temperature between θ and θ_c (possibly different in the two lines above). Since $\mathbf{A}(\theta^*), \mathbf{I}(\theta^*)$ are local minimizers of φ at θ^* , then the second terms within the large

parentheses vanish. If we assume that the entropy density ($= -\partial\varphi/\partial\theta$) is constant on the austenite and martensite wells, respectively, for temperatures in the interval (θ, θ_c) , identify the quantity

$$L = \theta_c \left(\frac{\partial\varphi(\mathbf{I}(\theta_c), \theta_c)}{\partial\theta} - \frac{\partial\varphi(\mathbf{A}(\theta_c), \theta_c)}{\partial\theta} \right) \quad (30)$$

as the latent heat, and use that the wells are at the same height at θ_c , we get by subtraction of (29)

$$\varphi(\mathbf{A}, \theta) - \varphi(\mathbf{I}, \theta) = L \frac{\theta_c - \theta}{\theta_c}. \quad (31)$$

We identify $(1/2)H = (\theta_c - \theta)$ as the half width of the hysteresis when (28) holds. This gives the criterion

$$\lambda \sqrt{\frac{2\kappa\mu(1-\lambda)\xi}{\ell_c}} = \varphi(\mathbf{A}, \theta) - \varphi(\mathbf{I}, \theta) = L \frac{\theta_c - \theta}{\theta_c} = \frac{1}{2}HL \frac{1}{\theta_c}, \quad (32)$$

that is, the width of the hysteresis is

$$H = \frac{2\lambda\theta_c}{L} \sqrt{\frac{2\kappa\mu(1-\lambda)\xi}{\ell_c}}. \quad (33)$$

To compare with Figure 3, the dependence of H on λ_2 near $\lambda_2 = 1$ is of interest. This arises from the dependence of λ on λ_2 in (33). This is given by $\lambda = 1 - \lambda^*$ with λ^* evaluated using (6) and (14). Its dominant term as λ_2 approaches 1 is

$$\begin{aligned} \lambda(\lambda_2) &= \frac{1}{2} - \frac{1}{2} \sqrt{\frac{-2|\lambda_2^2 - 1|}{\left| \frac{a_1 n_1 \lambda_1 (\lambda_2^2 - 1)}{\lambda_1^2 - 1} + \lambda_2 a_2 n_2 + \frac{a_3 n_3 \lambda_3 (\lambda_2^2 - 1)}{\lambda_3^2 - 1} \right|} + 1} \\ &\sim \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{4|\lambda_2 - 1|}{|a_2 n_2|}}. \end{aligned} \quad (34)$$

The slope at $\lambda_2 = 1$ is

$$\lambda'(\lambda_2) = \begin{cases} 1/|a_2 n_2|, & \lambda_2 = 1+, \\ -1/|a_2 n_2|, & \lambda_2 = 1-. \end{cases} \quad (35)$$

Putting (35) into (33) we get a formula for the asymptotic size of the hysteresis. This formula gives a qualitative prediction like that seen in the experimental data (Figure 3): two lines meeting at a sharp corner. Note that in Figure 5 the graph is not symmetric about $\lambda_2 = 1$. That is because both $a_2 n_2$ and ξ depend on the twin system which, as noted above, is different on the different sides of $\lambda_2 = 1$.

Note that in Table 3 the values of $a_2 n_2$ at $\lambda_2 = 1$ for a given twin system do not vary much among the alloys Ti-Ni-(Pd, Pt, Au, Cu-Pd). Similarly, the geometric factor *for a given twin system* also varies only slightly among these alloys. The other parameters $\mu, \theta_c, \kappa, L, \ell_c$ that enter the formula (33) for the hysteresis are also not expected to vary greatly among these alloys. Evidently, this insensitivity suggests a reason for the universality of the measured graph of hysteresis vs. λ_2 , Figure 3. According to the present theory, this universality is not expected to extend widely to other alloys:

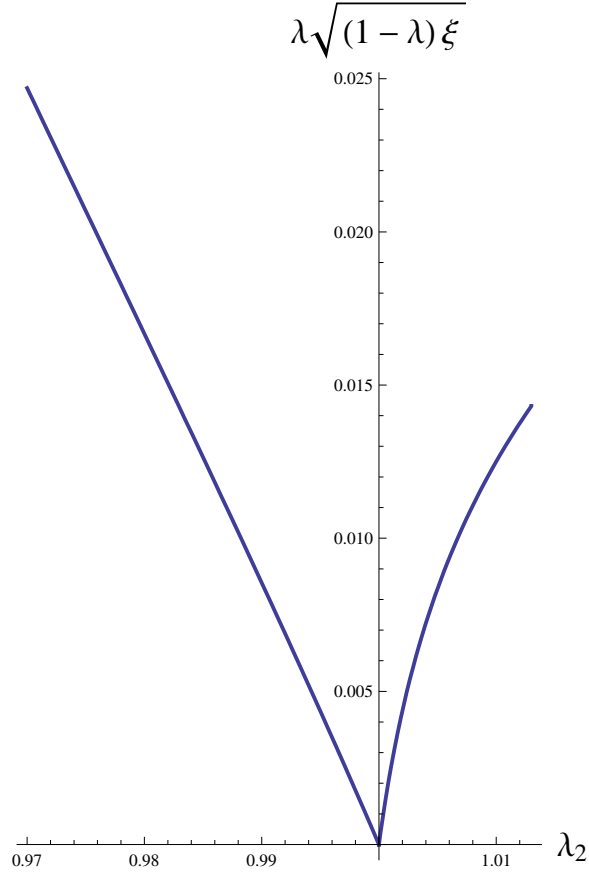


Figure 5: Plot of the factor in the formula for the hysteresis H vs. λ_2 that depends on lattice parameters. For this plot the other eigenvalues are determined by $\lambda_1 = 0.93$, $\lambda_1\lambda_2\lambda_3 = 1$ and the twin system is assumed to be the one with the lowest geometric factor ξ at the given λ_2 .

all of the parameters listed in this paragraph can be highly variable among different martensitic materials.

These results are also valid away from $\lambda_2 = 1$. Therefore it is interesting to plot the factor $\lambda\sqrt{(1-\lambda)\xi}$ in the formula for the hysteresis H vs. λ_2 . This is the factor that depends explicitly on lattice parameters. A plot of $\lambda\sqrt{(1-\lambda)\xi}$ vs. λ_2 is shown in Figure 5. Comparing Figure 5 with that of the data in Figure 3, one sees a remarkable similarity.

We cannot give more quantitative results because of the lack of knowledge of ℓ_c and of the interfacial energy constant κ . Regarding the latter, among all shape memory alloys, there are very few measured interfacial energies, those that are available are considered highly approximate by most workers. Schryvers [38] studies the twin branching phenomenon in NiAl. From the associated scaling law for twin spacing vs. distance from the habit plane he extracts a prefactor, and then, by use of an estimate of the elastic energy of the transition layer (similar to that done here) he gets an interfacial energy, roughly estimated at 1 mJ/m². For Type I twins in CuAlNi Shilo et al.[42] estimate 70 mJ/m² by fitting measured profiles of twin boundaries to a Landau theory. Waitz et al. [45] extracts atomic positions from high resolution images of twin boundaries and uses these to assign atomic positions for density functional theory (DFT) calculations, arriving at a value of 14

Table 5: Material constants used to estimate ℓ_c for $\text{Ti}_{50}\text{Ni}_{50-x}\text{Pd}_x$, $x = 11$

	material constant	reference
latent heat	$L = 0.231 \text{ J/g}$	Table 1, this paper
transformation temperature	$\theta_c = 298 \text{ K}$	Table 1, this paper
elastic modulus	$\mu = 27 \text{ GPa}$	ref. [33]
density	$\rho = 6.85 \text{ g/cm}^3$	Table 1, this paper
lattice parameters	$\lambda_1 = 0.9280, \lambda_2 = 1.0001, \lambda_3 = 1.0674$	Table 1, this paper
Interfacial energy density (compound twins)	$\kappa = 25 \text{ mJ/m}^2$	estimated, see text
$\lambda'(\lambda_2) _{\lambda_2=1^-} = -1/a_2 n_2$	-15.36	Table 3, this paper
$H'(\lambda_2) _{\lambda_2=1^-}$	-3000 K	Figure 3, this paper

mJ/m² for the compound twins and 187 mJ/m² for the Type I twins in monoclinic TiNi. It should be mentioned in this context that minimum energy calculations by Huang et al. [16] using several different implementations of DFT do not give the accepted stable phase of NiTi at $\theta = 0$ (i.e., the one seen in the high resolution images of Waitz et al. [45]). O’Handley et al. [32] give 40 mJ/m² for the compound twins in the tetragonal phase of an alloy of Ni₂MnGa.

Despite these uncertainties, it is useful to work backwards from the hysteresis data in Figure 3 to a value of the critical length ℓ_c . To do this, we use twin systems with the smallest geometric factor, as these will lose stability first according to the present viewpoint, and we consider the alloy $\text{Ti}_{50}\text{Ni}_{50-x}\text{Pd}_x$, $x \sim 11$. For this alloy, by direct calculation based on (22), the lowest geometric factors are obtained from the Type II twins of variants 1 and 4 (geometric factor 2.07×10^{-3}) and the compound twins of variants 1 and 2 (geometric factor 3.47×10^{-3}), and their symmetry related twin systems. The slopes of the of the measured hysteresis data can then be used, together with other assumed or measured material constants in Table 5, to give a value of ℓ_c using (33). The values we obtain by that procedure are $\ell_c = 17.5\mu\text{m}$ (compound) and $\ell_c = 10.4\mu\text{m}$ (Type II). Considering the uncertainty in the interfacial energy and other constants, these numbers could be easily off by factors of 2 or 3. However, they do suggest rather large values of ℓ_c which is consistent with the viewpoint adopted in this paper that the growth of fully developed austenite/martensite interfaces is responsible for determining the size of the hysteresis.

7 Remarks and comparisons with other ideas about hysteresis

It should be emphasized that the simple theory given above gives zero hysteresis when $\lambda_2 = 1$, whereas the measured thermal hysteresis decreased to between 5 and 15 C in the alloys studied. Possible reasons for this discrepancy are that 1) the alloys studied experimentally did not sufficiently closely satisfy $\lambda_2 = 1$ or were not sufficiently homogeneous, 2) the theory did not assign an energy to the exact interface between austenite and martensite in the case $\lambda_2 = 1$, and 3) other explanations besides the crossing of an energy barrier associated with a fully developed austenite/martensite interface, such as thermal activation (but see below) or pinning of interfaces by defects.

In the above we have attributed the crossing of the barrier to the presence of defects in the material, but a competing explanation is thermal activation, calculated, for example, by using transition state theory. In this context it is useful to compare the barrier height we have calculated to $k\theta$. For this purpose we need also values of other dimensions h, w of the martensite plate. In almost all observations of the first appearance of martensite, the plate is a thin plate, $h \gg \ell_c$, $w \gg \ell_c$. We then obtain a highly conservative estimate of the applicability of transition state theory by putting $h = w = \ell_c$. Using the smallest value $\ell_c = 10.4 \mu\text{m}$ calculated above, we get a barrier height of $E_b = 9.0 \times 10^{-11}$ J, while at $\theta = \theta_c$ we have $k\theta = 4.4 \times 10^{-21}$ J. Thus $E_b/k\theta_c \sim 2 \times 10^{10}$ and it is seen that the crossing of the energy barrier we have calculated here by a thermal fluctuation is exceedingly unlikely.

The latent heat L enters the formula (33) for the hysteresis in the denominator. In a sense θ_c/L sets the scale for conversion of energy to temperature difference. All else the same, materials with large latent heat will then have small apparent hysteresis. This is opposite to conventional wisdom, which holds that large latent heat implies large hysteresis because of the difficulty of getting the heat in and out of the material during transformation. The latter effect has been studied in detail by Leo and Shield [23] and Shaw and Kyriakides [41]. As demonstrated by these authors, this is a dynamic effect that has a negligible effect on the hysteresis loop with sufficiently slow transformation. Also, the relative changes of the size of the hysteresis due to this effect are rather small. In any case the formula (33) for H is not intended to capture this dynamic effect.

Materials with weak first order transformations often have relatively small hysteresis. In such materials, all the eigenvalues of the transformation strain matrix are close to 1, and in materials with small volume change, $\lambda_1\lambda_2\lambda_3 \sim 1$, λ_2 is particularly close to 1. In addition, the predicted asymptotic slopes of H vs. λ_2 at $\lambda_2 = 1$ tend to infinity as $\lambda_1, \lambda_3 \rightarrow 1$, because in that case $a_2n_2 \rightarrow 0$ (see (35)). Thus, the predicted graph of hysteresis H vs. λ_2 is sharper in the case of weak transformations, implying that one needs to approximate $\lambda_2 = 1$ with greater precision in order to achieve minimal hysteresis. However, κ and L are also expected to be small for such cases, so it is not clear how the hysteresis would be expected to behave for weak first order transformations. The asymptotic behavior of κ and L in cases where $\lambda_{1,2,3} \sim 1$ should be amenable to theoretical analysis and it would be interesting to make predictions about this case.

While this paper has concentrated on cubic to orthorhombic NiTiX alloys, there is significant evidence in the literature that the ideas presented here apply broadly. Some of this evidence, including comparisons with Cu-based martensitic materials undergoing cubic to monoclinic transformations, is given in [17]. The use of single crystals vs. polycrystals also does not affect greatly the hysteresis, also in accord with the theory given here. Qualitatively, the ideas appear to extend to stress-induced transformation. In principle, in the case of stress-induced transformation, one should have to account for the effect of stress on the local values of lattice parameters. The presence of a correlation between $\lambda_2 \sim 1$ and the size of the hysteresis in stress-induced cases is evidently a reflection of the smallness of the perturbation of the lattice parameters caused by the typical stresses used in such experiments. The effect of orientation on hysteresis seen in the experiments of Horikawa et al. [15] and of Shield [40] can be understood qualitatively [51] by using the measured elastic moduli to estimate the effect of stress on lattice parameters, so as to determine whether the stress moves the λ_2 closer to or further from the value 1.

One puzzling fact about the experimental results concerns the numbers of strains and numbers of variants. One would expect that these numbers might correlate with hysteresis, as large numbers of strains, for example, would seem to give more alternatives for accommodating stresses built up

in a polycrystal due to different grain orientations, and thereby would reduce or eliminate some energy barriers. But the data shown in Figure 3 is fairly symmetric about $\lambda_2 = 1$, even though the number of twins that participate in the austenite/martensite interface is drastically different for $\lambda_2 < 1$ or > 1 , as indicated in Section 6. The actual numbers of twin systems participating are 6 systems for $\lambda_2 < 1$ and 24 systems for $\lambda_2 > 1$. There are 4 austenite/martensite interfaces per twin system giving rise to 24 (resp., 96) interfaces for $\lambda_2 < 1$ (resp., > 1). Exactly at $\lambda_2 = 1$, where the hysteresis is lowest, these degenerate to only 12 austenite/martensite interfaces. Thus, surprisingly, there appears to be no obvious connection of the numbers of strains or numbers of interfaces to the size of hysteresis. This indicates again that the main energy barrier is the growth of an austenite/martensite interface, and the numbers of such interfaces is of secondary importance for the size of the hysteresis. However, these numbers could well play an important role for other properties, for example the lifetime of the material under repeated transformation. Interesting unpublished data of Quandt [36] supports this view. The satisfaction of the cofactor conditions described above achieves both a dramatically larger set of strains and interfaces, as well as $\lambda_2 = 1$.

8 Theory of the transition layer between austenite and martensite when $\lambda_2 \sim 1$

We now develop ideas for a transition layer that can provide a more accurate evaluation of the bulk energy of the transition layer than the one given above by simply using the matrix \mathbf{C} .

8.1 Kinematic assumptions and boundary conditions

We now formulate assumptions for a more general but still simplified transition layer. To begin, we look at common features of the deformation gradients that enter the boundary conditions for the transition layer. These are matrices $\mathbf{A}, \mathbf{B}, \mathbf{I}$ satisfying

$$\mathbf{B} - \mathbf{A} = \mathbf{a} \otimes \mathbf{n}, \quad (\lambda \mathbf{B} + (1 - \lambda) \mathbf{A}) - \mathbf{I} = \mathbf{b} \otimes \mathbf{m}. \quad (36)$$

Let \mathbf{e} be a unit vector perpendicular to both \mathbf{m} and \mathbf{n} , and let \mathbf{p} be a unit vector perpendicular to both \mathbf{a} and \mathbf{b} . From (36)

$$\mathbf{A}\mathbf{e} = \mathbf{B}\mathbf{e} = \mathbf{I}\mathbf{e} = \mathbf{e} \quad \text{and} \quad \mathbf{A}^T \mathbf{p} = \mathbf{B}^T \mathbf{p} = \mathbf{I}^T \mathbf{p} = \mathbf{p}. \quad (37)$$

Now we write the boundary conditions for a transition layer. On the left (see Figure 6) we have a martensite laminate. Its deformation can be written with the aid of the following periodic function with period 1:

$$\chi_\lambda(s) = \begin{cases} 1 - \lambda, & 0 \leq s < \lambda, \\ -\lambda, & \lambda \leq s < 1. \end{cases} \quad (38)$$

Note that χ_λ has mean 0, so its integral is also 1-periodic. The deformation of the martensite phase consistent with Figure 4 and the calculations above can be written

$$\mathbf{y}_M(\mathbf{x}) = (\lambda \mathbf{B} + (1 - \lambda) \mathbf{A}) \mathbf{x} + \varepsilon \mathbf{a} \int_0^{\frac{1}{\varepsilon} \mathbf{x} \cdot \mathbf{n}} \chi_\lambda(s) ds. \quad (39)$$

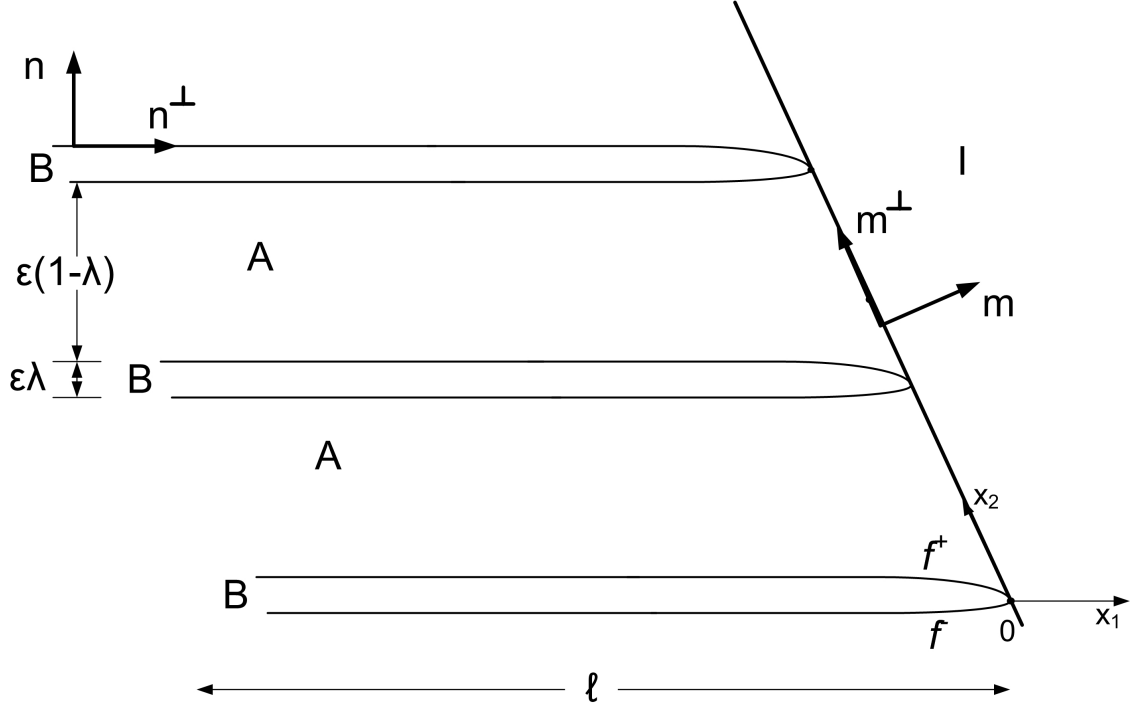


Figure 6: Austenite/martensite interface: notation for the Γ -convergence argument.

By direct calculation

$$\nabla \mathbf{y}_M(\mathbf{x}) = \begin{cases} \mathbf{B}, & \varepsilon i \leq \mathbf{x} \cdot \mathbf{n} < \varepsilon(i + \lambda), \\ \mathbf{A}, & \varepsilon(i + \lambda) \leq \mathbf{x} \cdot \mathbf{n} < \varepsilon(i + 1), \end{cases} \quad i \in \mathbb{Z}. \quad (40)$$

In the austenite we have $\nabla \mathbf{y}_A = \mathbf{I}$, so $\mathbf{y}_A(\mathbf{x}) = \mathbf{x} + \hat{\mathbf{c}}$. In principle, the presence of $\hat{\mathbf{c}}$ may be important; there should be a stiffness associated with changes of $\hat{\mathbf{c}}$. However, in the present treatment we shall assume (below) that the deformation gradient in the \mathbf{B} -layers is constant. This and the continuity of the deformation at 0 will then give a formula for $\hat{\mathbf{c}}$.

We write

$$\mathbf{y}(\mathbf{x}) = \mathbf{z}(\mathbf{x}) + \mathbf{x}. \quad (41)$$

We note from the expressions for \mathbf{y}_M and \mathbf{y}_A that $\nabla \mathbf{y}_M \mathbf{e} = \nabla \mathbf{y}_A \mathbf{e} = \mathbf{e}$. We assume that this property also extends to the transition layer:

$$\nabla \mathbf{z} \mathbf{e} = 0. \quad (42)$$

We can introduce coordinates

$$\mathbf{x} = x_1 \mathbf{n}^\perp + x_2 \mathbf{m}^\perp + x_3 \mathbf{e}. \quad (43)$$

Then we have by assumption $\mathbf{z}(x_1, x_2)$. Note that these coordinates are non-orthogonal (typically, $\mathbf{m}^\perp \cdot \mathbf{n}^\perp \neq 0$) but recall we have assumed $\mathbf{m}^\perp \cdot \mathbf{m}^\perp = \mathbf{n}^\perp \cdot \mathbf{n}^\perp = 1$ and $\mathbf{m}^\perp \cdot \mathbf{n} = \mathbf{m} \cdot \mathbf{n}^\perp$. Note also $\mathbf{x} \cdot \mathbf{n} = (\mathbf{n} \cdot \mathbf{m}^\perp)x_2$. As pictured in Figure 6, the austenite/martensite interface is $\mathbf{x} \cdot \mathbf{m} = 0$, that is, $x_1 = 0$. The twin boundaries are planes $x_2 = \text{const}$.

Now we discuss conditions of periodicity, first for the boundary conditions. We note that both $\mathbf{y}_A(\mathbf{x}) - \mathbf{x}$ and $\mathbf{y}_M(\mathbf{x}) - \mathbf{x}$ are periodic in the x_2 direction with period $\zeta = \varepsilon/(\mathbf{m}^\perp \cdot \mathbf{n})$. That's clear for $\mathbf{y}_A - \mathbf{x}$. For \mathbf{y}_M we have, using the second of (36),

$$\begin{aligned} \mathbf{y}_M(\mathbf{x} + \zeta \mathbf{m}^\perp) - \mathbf{y}_M(\mathbf{x}) - \zeta \mathbf{m}^\perp &= (\mathbf{I} + \mathbf{b} \otimes \mathbf{m})(\zeta \mathbf{m}^\perp) + \varepsilon \mathbf{a} \int_{\frac{1}{\varepsilon} \mathbf{x} \cdot \mathbf{n}}^{\frac{1}{\varepsilon}(\mathbf{x} + \zeta \mathbf{m}^\perp) \cdot \mathbf{n}} \chi_\lambda(s) ds - \zeta \mathbf{m}^\perp, \\ &= \varepsilon \mathbf{a} \int_{\frac{1}{\varepsilon} \mathbf{x} \cdot \mathbf{n}}^{\frac{1}{\varepsilon}(\mathbf{x} \cdot \mathbf{n}) + 1} \chi_\lambda(s) ds = 0. \end{aligned} \quad (44)$$

Consistent with this, note that the construction of the transition layer using \mathbf{C} also had the property that $\nabla \mathbf{y}$ is periodic in the x_2 direction with period $\zeta = \varepsilon/(\mathbf{m}^\perp \cdot \mathbf{n})$. Extending this hypothesis to more general transition layers, we assume that

$$\mathbf{z}(x_1, x_2 + \zeta) = \mathbf{z}(x_1, x_2), \quad \zeta = \frac{\varepsilon}{\mathbf{m}^\perp \cdot \mathbf{n}}. \quad (45)$$

We do not expect the twin boundaries necessarily to remain planar up to the austenite/martensite interface so now we parameterize these boundaries. We assume a \mathbf{B} -layer is confined by the two surfaces (see Figure 6)

$$\begin{aligned} x_2 &= f^+(x_1), \quad x_2 = f^-(x_1), \quad x_1 \leq 0, \\ f^+ &\geq f^-, \quad f^+(0) = f^-(0) = 0. \end{aligned} \quad (46)$$

By assuming that $f^+(0) = f^-(0) = 0$ we have put the origin in its natural place at the intersection of the pinched needle and the austenite/martensite interface. The needle is assumed to be pinched because it is the \mathbf{A} -layer that is nearly compatible with austenite when $\lambda_2 \sim 1$. We do not necessarily want to assume the needle is symmetric, or even that the centerline of the \mathbf{B} -layer (as obtained from the boundary conditions) passes through the origin. Thus, we introduce a parameter η and write the boundary conditions for a martensite plate of length ℓ as the following:

$$\begin{aligned} \mathbf{y}(\mathbf{x}) &= \mathbf{y}_M(\mathbf{x} - \eta \varepsilon \mathbf{n}) \text{ at } x_1 = -\ell, \text{ where} \\ f^-(-\ell) &= \eta \zeta \quad \text{and} \quad f^+(-\ell) = (\eta + \lambda) \zeta, \end{aligned} \quad (47)$$

(Note for example that the \mathbf{B} -layer $\eta \varepsilon < \mathbf{x} \cdot \mathbf{n} < (\eta + \lambda) \varepsilon$ corresponds to $\eta \zeta < x_2 < (\eta + \lambda) \zeta$.) The parameter η is also part of the solution; we expect it will emerge that $\eta < 0$. The displacement \mathbf{z} then satisfies the boundary condition

$$\mathbf{z}(\mathbf{x}) = \mathbf{y}_M(\mathbf{x} - \eta \varepsilon \mathbf{n}) - \mathbf{x}, \quad x_1 = -\ell, \quad (48)$$

In (47) and (48) we have imposed the left hand boundary condition of the \mathbf{A} -layer at $x_1 = -\ell$. Typically, we will want $\ell \gg \varepsilon$ so as to capture the full transition layer.

For simplicity we assume constant gradient in the \mathbf{B} -layers:

$$\begin{aligned} \mathbf{z}(\mathbf{x}) &= \mathbf{y}_M(\mathbf{x} - \eta \varepsilon \mathbf{n}) - \mathbf{x} = (\mathbf{B} - \mathbf{I})\mathbf{x} - \eta \varepsilon \mathbf{B}\mathbf{n} = x_1(\mathbf{B} - \mathbf{I})\mathbf{n}^\perp + x_2(\mathbf{B} - \mathbf{I})\mathbf{m}^\perp - \eta \varepsilon \mathbf{B}\mathbf{n} \\ &\text{for } f^-(x_1) < x_2 < f^+(x_1). \end{aligned} \quad (49)$$

This defines \mathbf{z} on the \mathbf{B} -layer attached to the origin. The periodicity condition $\mathbf{z}(x_1, x_2 + \zeta) = \mathbf{z}(x_1, x_2)$ extends this function to the other \mathbf{B} -layers. For the \mathbf{A} -layer $f^+(x_1) < x_2 < f^-(x_1) + \zeta$ we therefore have the boundary conditions,

$$\begin{aligned} \mathbf{z}(x_1, f^+(x_1)) &= x_1(\mathbf{B} - \mathbf{I})\mathbf{n}^\perp + f^+(x_1)(\mathbf{B} - \mathbf{I})\mathbf{m}^\perp - \eta\varepsilon\mathbf{Bn}, \\ \mathbf{z}(x_1, f^-(x_1) + \zeta) &= x_1(\mathbf{B} - \mathbf{I})\mathbf{n}^\perp + f^-(x_1)(\mathbf{B} - \mathbf{I})\mathbf{m}^\perp - \eta\varepsilon\mathbf{Bn}, \end{aligned} \quad (50)$$

for $x_1 < 0$, and, on this layer ($f^+(x_1) < x_2 < f^-(x_1) + \zeta$) we expect $\nabla\mathbf{y} \sim \text{SO}(3)\mathbf{A}$. Making the ansatz (49) for the \mathbf{B} -layers now gives us a way to evaluate the constant $\hat{\mathbf{c}}$ by continuity at $\mathbf{x} = 0$:

$$\mathbf{y}(0) = \mathbf{z}(0) = \hat{\mathbf{c}} = -\eta\varepsilon\mathbf{Bn}. \quad (51)$$

At the sharp interface with austenite as pictured in Figure 6 we also have the boundary condition

$$\mathbf{z}(0, x_2) = (\mathbf{y}_A(\mathbf{x}) - \mathbf{x})|_{x_1=0} = -\eta\varepsilon\mathbf{Bn}, \quad 0 < x_2 < \zeta, \quad (52)$$

in view of (51). Finally, we have the conditions at $x_1 = -\infty$ for the \mathbf{A} -layer:

$$\begin{aligned} \mathbf{z}(-\ell, x_2) = \mathbf{y}_M(\mathbf{x} - \eta\varepsilon\mathbf{n}) - \mathbf{x} &= (\mathbf{A} - \mathbf{I})\mathbf{x} + \varepsilon\lambda\mathbf{a} - \eta\varepsilon\mathbf{An} \quad \text{at } x_1 = -\ell, \\ &= x_1(\mathbf{A} - \mathbf{I})\mathbf{n}^\perp + x_2(\mathbf{A} - \mathbf{I})\mathbf{m}^\perp + \varepsilon\lambda\mathbf{a} - \eta\varepsilon\mathbf{An} \end{aligned} \quad (53)$$

for $(\eta + \lambda)\zeta < x_2 < (\eta + 1)\zeta$. A summary of boundary conditions for the \mathbf{A} -layer and the conditions on f^\pm is given below.

Boundary conditions on the \mathbf{A} -layer

$$\begin{aligned} f^-(-\ell) &= \eta\zeta \quad \text{and} \quad f^+(-\ell) = (\eta + \lambda)\zeta, \quad \zeta = \frac{\varepsilon}{\mathbf{m}^\perp \cdot \mathbf{n}}, \\ \mathbf{z}(x_1, f^+(x_1)) &= x_1(\mathbf{B} - \mathbf{I})\mathbf{n}^\perp + f^+(x_1)(\mathbf{B} - \mathbf{I})\mathbf{m}^\perp - \eta\varepsilon\mathbf{Bn}, \quad x_1 < 0, \\ \mathbf{z}(x_1, f^-(x_1) + \zeta) &= x_1(\mathbf{B} - \mathbf{I})\mathbf{n}^\perp + f^-(x_1)(\mathbf{B} - \mathbf{I})\mathbf{m}^\perp - \eta\varepsilon\mathbf{Bn}, \quad x_1 < 0, \\ \mathbf{z}(0, x_2) &= -\eta\varepsilon\mathbf{Bn}, \quad 0 < x_2 < \zeta, \\ \mathbf{z}(-\ell, x_2) &= -\ell(\mathbf{A} - \mathbf{I})\mathbf{n}^\perp + x_2(\mathbf{A} - \mathbf{I})\mathbf{m}^\perp + \varepsilon\lambda\mathbf{a} - \eta\varepsilon\mathbf{An}, \\ &\quad (\eta + \lambda)\zeta < x_2 < (\eta + 1)\zeta. \end{aligned} \quad (54)$$

8.2 Change of variables

We now make a change of variables that brings out the structure of the basic mathematical problem and removes the solution-dependence of the domain of integration for the energy. The basic small parameter for this problem can be taken as either λ or $\lambda_2 - 1$, these being related by (34).

For the independent variables we change variables,

$$t_1 = x_1, \quad t_2 = \frac{x_2 - f^+(x_1)}{\zeta - (f^+(x_1) - f^-(x_1))}. \quad (55)$$

The Jacobian of (the inverse of) this transformation is

$$J = \frac{\partial(x_1, x_2)}{\partial(t_1, t_2)} = \zeta - (f^+ - f^-). \quad (56)$$

This transformation maps the domain $\{-\infty < x_1 < 0, f^+(x_1) < x_2 < f^-(x_1) + \zeta\}$ to the fixed half-strip $\{-\infty < t_1 < 0, 0 < t_2 < 1\}$: the \mathbf{A} -layer becomes a rectangular half-strip in these new variables. For the dependent variables we make the change to

$$\mathbf{u}(t_1, t_2) = \frac{1}{\mathbf{n}^\perp \cdot \mathbf{m}} \left(\mathbf{z}(t_1, [\zeta - (f^+(t_1) - f^-(t_1))]t_2 + f^+(t_1)) - t_1(\mathbf{A} - \mathbf{I})\mathbf{n}^\perp + \eta\varepsilon\mathbf{B}\mathbf{n} \right). \quad (57)$$

Recalling that $\mathbf{y}(\mathbf{x}) = \mathbf{z}(\mathbf{x}) + \mathbf{x}$ and $x_1 = \mathbf{x} \cdot \mathbf{m}/(\mathbf{n}^\perp \cdot \mathbf{m})$, $x_2 = \mathbf{x} \cdot \mathbf{n}/(\mathbf{n}^\perp \cdot \mathbf{m})$, (cf. (43)), we denote

$$g = \frac{t_2 f'^- + (1 - t_2) f'^+}{\zeta - (f^+ - f^-)}, \quad h = \frac{1}{\zeta - (f^+ - f^-)}, \quad (58)$$

and then we have,

$$\begin{aligned} \nabla \mathbf{y} &= \mathbf{I} + \frac{1}{\mathbf{n}^\perp \cdot \mathbf{m}} (\mathbf{z}_{,1} \otimes \mathbf{m} + \mathbf{z}_{,2} \otimes \mathbf{n}) \\ &= \mathbf{I} + \left(\mathbf{u}_{,1} + \frac{1}{\mathbf{n}^\perp \cdot \mathbf{m}} (\mathbf{A} - \mathbf{I})\mathbf{n}^\perp - g\mathbf{u}_{,2} \right) \otimes \mathbf{m} + h\mathbf{u}_{,2} \otimes \mathbf{n} \\ &= \mathbf{I} + \frac{1}{\mathbf{n}^\perp \cdot \mathbf{m}} (\mathbf{A} - \mathbf{I})\mathbf{n}^\perp \otimes \mathbf{m} + (\mathbf{u}_{,1} - g\mathbf{u}_{,2}) \otimes \mathbf{m} + h\mathbf{u}_{,2} \otimes \mathbf{n}. \end{aligned} \quad (59)$$

From the equations of the crystallographic theory we have $\lambda(\mathbf{B} - \mathbf{A}) + \mathbf{A} - \mathbf{I} = \mathbf{b} \otimes \mathbf{m}$, implying that $(\mathbf{A} - \mathbf{I})\mathbf{n}^\perp = (\mathbf{m} \cdot \mathbf{n}^\perp)\mathbf{b}$. Substituting this into the second term on the right hand side of (59), and then using that $\mathbf{I} + \mathbf{b} \otimes \mathbf{m} = \mathbf{A} + \lambda\mathbf{a} \otimes \mathbf{n}$ we get the convenient form

$$\nabla \mathbf{y} = \mathbf{A} + (\mathbf{u}_{,1} - g\mathbf{u}_{,2}) \otimes \mathbf{m} + (\lambda\mathbf{a} + h\mathbf{u}_{,2}) \otimes \mathbf{n}. \quad (60)$$

Finally, directly from (54), we have the boundary conditions for the function \mathbf{u} .

Boundary conditions for $\mathbf{u}(t_1, t_2)$

$$\begin{aligned} f^-(-\ell) &= \eta\zeta \quad \text{and} \quad f^+(-\ell) = (\eta + \lambda)\zeta, \quad \zeta = \frac{\varepsilon}{\mathbf{m}^\perp \cdot \mathbf{n}}, \\ \mathbf{u}(t_1, 0) &= f^+(t_1)(1 - \lambda)\mathbf{a}, \quad t_1 < 0, \\ \mathbf{u}(t_1, 1) &= f^-(t_1)(1 - \lambda)\mathbf{a}, \quad t_1 < 0, \\ \mathbf{u}(0, t_2) &= 0, \quad 0 < t_2 < 1, \\ \mathbf{u}(-\ell, t_2) &= \zeta((\lambda + \eta) - \lambda t_2)(1 - \lambda)\mathbf{a}, \quad 0 < t_2 < 1. \end{aligned} \quad (61)$$

8.3 Scaling of the energy

At this point we could do a series of numerical energy minimizations to determine the energy in the transition layer as a function of the parameters. We want to vary at least the small parameter λ , the elastic modulus μ and the other two eigenvalues λ_1, λ_3 in the transformation strain matrix. (The effect of varying the fineness ε , or, equivalently, the periodicity ζ , is determined by the basic scale-invariance of elasticity). We may also wish to consider more general elastic energy functions. Thus, a comprehensive series of numerical minimizations like this could quickly become unwieldy, and it is not clear whether a simple understanding of the dependence of the parameters would be

emerge. This is a good situation for Γ -convergence, as this method gives the leading order energy, the leading order minimizer, and simple predictions of the effect of changing other parameters.

To begin, we first need to understand the basic scaling of the energy with λ . For this purpose the earlier test function involving the triangles with deformation gradient \mathbf{C} is expected to give the correct scaling, though not the optimal energy. This test function is easily expressed in the simplified variables of the preceding subsection. It corresponds to the choice

$$\eta = 0, \quad f^-(t_1) = 0, \quad f^+(t_1) = \begin{cases} \lambda\zeta, & -\ell < t_1 < \hat{\alpha}, \\ \frac{\lambda\zeta}{\hat{\alpha}}t_1, & \hat{\alpha} \leq t_1 < 0, \end{cases} \quad (62)$$

where $\hat{\alpha} = \alpha/(\mathbf{n}^\perp \cdot \mathbf{m})$, with $\alpha < 0$ having the same meaning as in Section 5. The function \mathbf{u} is chosen as the simple continuous function

$$\mathbf{u}(t_1, t_2) = \begin{cases} \zeta\lambda(1-t_2)(1-\lambda)\mathbf{a}, & -\ell < t_1 < \hat{\alpha}, \quad 0 < t_2 < 1, \\ \zeta\lambda(1-t_2)(1-(\lambda/\hat{\alpha})t_1)\mathbf{a}, & \hat{\alpha} \leq t_1 < \hat{\alpha}(t_2-1)/(\lambda t_2-1), \quad 0 < t_2 < 1, \\ (\zeta\lambda t_1(1-\lambda)/\hat{\alpha})\mathbf{a}, & \hat{\alpha}(t_2-1)/(\lambda t_2-1) < t_1 < 0, \quad 0 < t_2 < 1. \end{cases} \quad (63)$$

With these choices one can see that all the boundary conditions (61) for \mathbf{u} are satisfied, and, using the expression (60) we find that the deformation gradient on the region $\hat{\alpha}(t_2-1)/(\lambda t_2-1) < t_1 < 0$, is

$$\nabla \mathbf{y} = \mathbf{A} + \mathbf{u}_{,1} \otimes \mathbf{m} + \lambda \mathbf{a} \otimes \mathbf{n} = \mathbf{A} + \mathbf{a} \otimes \left(\frac{\varepsilon\lambda(1-\lambda)}{\alpha} \mathbf{m} + \lambda \mathbf{n} \right) = \mathbf{C}, \quad (64)$$

(cf., (17)), and of course outside this region $\nabla \mathbf{y} = \mathbf{A}$. This is precisely the test function used earlier in this paper, expressed in the new variables. Therefore we can rely on the earlier evaluation (19) and we see immediately that the energy goes as λ^2 .

Beginning with a general expression for the energy, we therefore divide the energy by λ^2 and seek the Γ limit of

$$\frac{1}{\lambda^2} \int_{-\ell}^0 \int_0^1 \tilde{\varphi}(\mathbf{I} + (\mathbf{u}_{,1} - g\mathbf{u}_{,2}) \otimes \mathbf{A}^{-T} \mathbf{m} + (\lambda \mathbf{a} + h\mathbf{u}_{,2}) \otimes \mathbf{A}^{-T} \mathbf{n}) (\zeta - (f^+ - f^-)) dt_1 dt_2, \quad (65)$$

where $\tilde{\varphi}$ is minimized on $\text{SO}(3)$, that is, $\tilde{\varphi}(\mathbf{F}) = \varphi(\mathbf{F}\mathbf{A}, \theta)$ where φ is the free energy used earlier in this paper and the dependence on θ is suppressed.

By dividing by λ^2 we are assured that the rescaled energy of the test function used earlier in this paper is bounded away from 0 and ∞ as $\lambda \rightarrow 0$ (cf., (21)). But of course, this test function is not a minimizer of the energy. Note also that the integrand of (65) is minimized when $\mathbf{u}_{,1} = 0$, $\mathbf{u}_{,2} = -\zeta\lambda(1-\lambda)\mathbf{a}$, $f^- = \eta\zeta$, $f^+ = (\eta + \lambda)\zeta$ but of course this satisfies only some of the boundary conditions. In particular it does not satisfy the boundary condition at $t_1 = 0$.

8.4 Calculation of the Γ -limit of the energy

8.4.1 Hypotheses and summary of results used in the argument

In this section we collect the hypotheses used for the derivation of the Γ -limit and, for the convenience of the reader, we collect the various results and definitions used in the argument. The main hypothesis is a bound on the slopes of the functions f^\pm that has the effect of preventing very thin fingers of the \mathbf{A} -layer from penetrating into the \mathbf{B} -layer. This assumption could be related to

the possibility of formation of nonclassical austenite/martensite interfaces [1], a possibility that is therefore being ruled out. A second main hypothesis says that $\tilde{\varphi}$ grows at least quadratically as we move away from its energy well, $\text{SO}(3)$.

Hypotheses 1) We assume $f^+ < f^- + \zeta$ and f^\pm are Lipschitz functions with Lipschitz constant $c > 0$. The latter means that $|f^\pm(s_2) - f^\pm(s_1)| \leq c|s_2 - s_1|$ for all $-\ell < s_1 < s_2 < 0$. It also implies that the region $\Omega_\ell = \{(x_1, x_2) : f^+(x_1) < x_2 < f^-(x_1) + \zeta, -\ell < x_1 < 0\}$ (called here the **A**-region) is a Lipschitz domain, and therefore the rigidity lemma (see below) can be applied on Ω_ℓ . 2) The energy density $\tilde{\varphi}$ is assumed to be twice differentiable and to satisfy $\tilde{\varphi}(\mathbf{F}) \geq C \text{dist}^2(\mathbf{F}, \text{SO}(3))$. 3) We assume a solution $\mathbf{A} = \mathbf{A}_\lambda$ and $\mathbf{B} = \mathbf{B}_\lambda$ and $\mathbf{b} = \mathbf{b}_\lambda$, $\mathbf{m} = \mathbf{m}_\lambda$, $\mathbf{m}_\lambda \cdot \mathbf{n}^\perp = 1$, of the crystallographic theory (4) with $\mathbf{A}_\lambda \rightarrow \mathbf{A}_0 \in \text{SO}(3)\mathbf{U}_1|_{\lambda_2=1} \neq \mathbf{I}$ as $\lambda \rightarrow 0$, and $\mathbf{A}_\lambda - \mathbf{B}_\lambda = \mathbf{a}_\lambda \otimes \mathbf{n}$, $\mathbf{a}_0 \neq 0$.

Summary of definitions and results used 1) The notation $\text{dist}(\mathbf{F}, \mathcal{S})$ denotes the shortest Euclidean distance from the 3×3 matrix \mathbf{F} to the (closed, bounded) set \mathcal{S} . In the case rotation matrices, $\mathcal{S} = \text{SO}(3) = \{\mathbf{R} : \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det \mathbf{R} = 1\}$, and assuming $\det \mathbf{F} > 0$, an alternative expression is $\text{dist}(\mathbf{F}, \text{SO}(3)) = |\mathbf{U} - \mathbf{I}|$ where \mathbf{U} is the right stretch tensor of elasticity theory, i.e., $\mathbf{F} = \mathbf{R}\mathbf{U}$ is the polar decomposition of \mathbf{F} . The notations L^∞, L^2, H^1 denote spaces of functions defined, respectively, as the set of functions for which the the following norms are finite: for $L^\infty(\Omega)$, $\|f\|_{L^\infty(\Omega)} = \text{ess sup}_\Omega |f|$; for $L^2(\Omega)$, $\|f\|_{L^2(\Omega)}^2 = \int_\Omega |f|^2 d\mathbf{x}$; for $H^1(\Omega)$, $\|f\|_{H^1(\Omega)}^2 = \int_\Omega |f|^2 d\mathbf{x} + \int_\Omega |\nabla f|^2 d\mathbf{x}$. **Something about weak convergence here, used in L^2 and H^1 so far.** The interpolation inequality used below states that for any Lipschitz function f , we have $\|f\|_{L^\infty(\mathbb{R})} \leq C \|f\|_{L^2(\mathbb{R})}^{\frac{2}{3}} \|f'\|_{L^\infty(\mathbb{R})}^{\frac{1}{3}}$. The following Rigidity Lemma [12] will also be used in the argument: for each given function $\mathbf{y} \in H^1(\Omega)$ there is a corresponding rotation matrix $\bar{\mathbf{R}} \in \text{SO}(3)$ such that

$$\int_\Omega |\nabla \mathbf{y}(\mathbf{x}) - \bar{\mathbf{R}}|^2 d\mathbf{x} \leq C \int_\Omega \text{dist}^2(\nabla \mathbf{y}(\mathbf{x}), \text{SO}(3)) d\mathbf{x}. \quad (66)$$

Here C depends only on $\Omega \in \mathbb{R}^3$, which is assumed to be a Lipschitz domain. This is a quantitative statement of the fact that, if a deformation gradient is at each $\mathbf{x} \in \Omega$ near some rotation matrix, which *a priori* could vary from point to point in Ω , it is actually near a single rotation matrix on most of Ω .

Below, the “original variables” refers to deformations expressed as a function of x_1, x_2, x_3 and satisfying boundary conditions (54) while “new variables” refers to functions \mathbf{u} (see (57)) expressed as a function of t_1, t_2, t_3 and satisfying the equivalent boundary conditions (61).

8.4.2 Use of the rigidity lemma to restrict the form of a minimizer

Suppose $\mathbf{y}_\lambda, f_\lambda^\pm$ have less energy than the naive test function (63) and satisfy the boundary conditions (54). Then $\mathbf{y}_\lambda, f_\lambda^\pm$ satisfy

$$\left(\int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} \text{dist}^2(\nabla \mathbf{y}_\lambda, \text{SO}(3)\mathbf{A}_\lambda) dx_2 dx_1 \leq \right) \int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} \varphi(\nabla \mathbf{y}_\lambda, \theta) dx_2 dx_1 \leq C\lambda^2, \quad (67)$$

the lower bound in parentheses being part of our hypotheses above. We first use of the rigidity lemma in the original variables to restrict the rotation. We note that the rigidity lemma (66) remains true also with $\bar{\mathbf{R}}$ replaced by $\mathbf{R}\mathbf{A}_\lambda$ and $\text{SO}(3)$ concurrently replaced by $\text{SO}(3)\mathbf{A}_\lambda$, by a

linear change of variables. Applying this lemma we therefore have the existence of $\mathbf{R}_\lambda \in \text{SO}(3)$ such that

$$\int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} |\nabla \mathbf{y}_\lambda - \mathbf{R}_\lambda \mathbf{A}_\lambda|^2 dx_2 dx_1 \leq C\lambda^2. \quad (68)$$

By passing to a subsequence we can assume $\mathbf{R}_\lambda \rightarrow \mathbf{R}_0$. In view of the uniform Lipschitz bound, $f_\lambda^\pm \rightarrow f^\pm$ uniformly. Thus, we can assert (68) on a fixed domain, i.e.,

$$\int_{-\ell}^0 \int_{f^++\delta}^{\zeta+f^--\delta} |\nabla \mathbf{y}_\lambda - \mathbf{R}_\lambda \mathbf{A}_\lambda|^2 dx_2 dx_1 \leq C\lambda^2. \quad (69)$$

for $\delta > 0$. Then we use the trace theorem applied to the sub-boundary $\mathcal{S} = \{(x_1, x_2) : x_1 = 0, f^+(0) + \delta < x_2 < \zeta + f^-(0) - \delta\}$ to deduce that there is a constant \mathbf{c}_λ such that

$$\int_{\mathcal{S}} |\mathbf{y}_\lambda - \mathbf{R}_\lambda \mathbf{A}_\lambda \mathbf{x} - \mathbf{c}_\lambda|^2 dx_2 \leq C\lambda^2. \quad (70)$$

Let $\mathbf{x}_0^\lambda \cdot \mathbf{m}_\lambda = 0$, $\mathbf{x}_0^\lambda \cdot \mathbf{n} = (1/2)(f^+ + f^- + \zeta)\mathbf{n} \cdot \mathbf{m}^\perp$ and let $B(\mathbf{x}_0^\lambda, r)$ be a ball of radius r centered at \mathbf{x}_0^λ . Written without coordinates, with the boundary condition $\mathbf{y}_\lambda(\mathbf{x}) = \mathbf{x}$ inserted, and restricted to a subdomain, (70) is

$$\int_{\{\mathbf{x} \cdot \mathbf{m}_\lambda = 0\} \cap B(\mathbf{x}_0^\lambda, r)} |(\mathbf{I} - \mathbf{R}_\lambda \mathbf{A}_\lambda) \mathbf{x} - \mathbf{c}_\lambda|^2 dA \leq C\lambda^2. \quad (71)$$

Alternatively, putting $\mathbf{F}_\lambda = \mathbf{I} - \mathbf{R}_\lambda \mathbf{A}_\lambda$,

$$\int_{\{\mathbf{x} \cdot \mathbf{m}_\lambda = 0\} \cap B(\mathbf{x}_0^\lambda, r)} |\mathbf{F}_\lambda(\mathbf{x} - \mathbf{x}_0^\lambda) - \mathbf{c}_\lambda + \mathbf{F}_\lambda \mathbf{x}_0^\lambda|^2 dA \leq C\lambda^2. \quad (72)$$

Expanding the integrand and using that the domain of integration is symmetric about \mathbf{x}_0^λ we get

$$\begin{aligned} & \int_{\{\mathbf{x} \cdot \mathbf{m}_\lambda = 0\} \cap B(\mathbf{x}_0^\lambda, r)} |(\mathbf{I} - \mathbf{R}_\lambda \mathbf{A}_\lambda)(\mathbf{x} - \mathbf{x}_0^\lambda) - \mathbf{c}_\lambda + (\mathbf{I} - \mathbf{R}_\lambda \mathbf{A}_\lambda) \mathbf{x}_0^\lambda|^2 dA \\ &= (\mathbf{F}_\lambda^T \mathbf{F}_\lambda) \cdot \int_{\{\mathbf{x} \cdot \mathbf{m}_\lambda = 0\} \cap B(\mathbf{x}_0^\lambda, r)} (\mathbf{x} - \mathbf{x}_0^\lambda) \otimes (\mathbf{x} - \mathbf{x}_0^\lambda) d\mathbf{x} + \pi r^2 |\mathbf{c}_\lambda - \mathbf{F}_\lambda \mathbf{x}_0^\lambda|^2 \leq C\lambda^2. \end{aligned} \quad (73)$$

Hence, the two positive semi-definite terms must satisfy,

$$|\mathbf{c}_\lambda - \mathbf{F}_\lambda \mathbf{x}_0^\lambda|^2 < C\lambda^2 \quad \text{and} \quad (\mathbf{F}_\lambda^T \mathbf{F}_\lambda) \cdot (\mathbf{I} - \mathbf{m}_\lambda \otimes \mathbf{m}_\lambda) < C\lambda^2. \quad (74)$$

We focus on the second of these and recall from the crystallographic theory that $\mathbf{A}_\lambda = \mathbf{I} - \lambda \mathbf{a}_\lambda \otimes \mathbf{n} + \mathbf{b}_\lambda \otimes \mathbf{m}_\lambda$. Using that $(\mathbf{I} - \mathbf{m}_\lambda \otimes \mathbf{m}_\lambda) \mathbf{m}_\lambda = 0$ we have that

$$(\mathbf{F}_\lambda^T \mathbf{F}_\lambda) \cdot (\mathbf{I} - \mathbf{m}_\lambda \otimes \mathbf{m}_\lambda) = (\mathbf{I} - \mathbf{R}_\lambda - \lambda \mathbf{R}_\lambda \mathbf{a}_\lambda \otimes \mathbf{n})^T (\mathbf{I} - \mathbf{R}_\lambda - \lambda \mathbf{R}_\lambda \mathbf{a}_\lambda \otimes \mathbf{n}) \cdot (\mathbf{I} - \mathbf{m}_\lambda \otimes \mathbf{m}_\lambda) \leq C\lambda^2. \quad (75)$$

This yields

$$(\mathbf{I} - \mathbf{R}_\lambda)^T (\mathbf{I} - \mathbf{R}_\lambda) \cdot (\mathbf{I} - \mathbf{m}_\lambda \otimes \mathbf{m}_\lambda) \leq C\lambda^2 + 2\lambda [(\mathbf{R}_\lambda - \mathbf{I}) \mathbf{a}_\lambda \cdot \mathbf{n} - (\mathbf{m}_\lambda \cdot (\mathbf{R}_\lambda - \mathbf{I}) \mathbf{a}_\lambda) (\mathbf{n} \cdot \mathbf{m}_\lambda)]. \quad (76)$$

Noting that $\max_{|\mathbf{s}|=1} |(\mathbf{R}_\lambda - \mathbf{I})\mathbf{s}|^2 = 2(1 - \cos \theta_\lambda) = (1/2)|\mathbf{R}_\lambda - \mathbf{I}|^2$, we have first from (76) that $|\mathbf{R}_\lambda - \mathbf{I}|^2 \leq C\lambda$ and then, again from (76), that

$$|\mathbf{R}_\lambda - \mathbf{I}| \leq C\lambda. \quad (77)$$

Using again the form of \mathbf{A}_λ and the triangle inequality,

$$\|\nabla \mathbf{y}_\lambda - \mathbf{A}_0\|_{L^2} \leq \|\nabla \mathbf{y}_\lambda - \mathbf{R}_\lambda \mathbf{A}_\lambda\|_{L^2} + \|(\mathbf{R}_\lambda - \mathbf{I})\mathbf{A}_\lambda\|_{L^2} + \|\mathbf{A}_\lambda - \mathbf{A}_0\|_{L^2}, \quad (78)$$

we have that

$$\int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} |\nabla \mathbf{y}_\lambda - \mathbf{A}_0|^2 dx_2 dx_1 \leq C\lambda^2, \quad (79)$$

(or, the same bound with \mathbf{A}_0 replaced by \mathbf{A}_λ). In summary, the rigidity lemma implies that a deformation with energy less than $\sim \lambda^2$ is actually be close to the matrix \mathbf{A}_0 in L^2 , the quantitative bound being given by (79).

8.4.3 Bounds on the positions f^\pm of the twin boundaries arising from energy minimization

We have assumed that $0 \leq f^+ - f^- \leq \zeta$. Staying for now with the original variables we have

$$\int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} |(\nabla \mathbf{y}_\lambda - \mathbf{A}_\lambda)\mathbf{m}^\perp|^2 dx_2 dx_1 \leq \int_{-\ell}^0 \int_{f_\lambda^+}^{\zeta+f_\lambda^-} |\nabla \mathbf{y}_\lambda - \mathbf{A}_\lambda|^2 dx_2 dx_1 \leq C\lambda^2. \quad (80)$$

(Recall that $\nabla \mathbf{y} = (1/\mathbf{m}^\perp \cdot \mathbf{n})(\mathbf{y}_{,1} \otimes \mathbf{m} + \mathbf{y}_{,2} \otimes \mathbf{n}) + \mathbf{y}_{,3} \otimes \mathbf{e}$. An application of Jensen's inequality ($\int_a^b f^2 dx \geq (1/(b-a))(\int_a^b f dx)^2$) to the inner integral gives

$$\int_{-\ell}^0 \left| \int_{f_\lambda^+}^{\zeta+f_\lambda^-} (\nabla \mathbf{y}_\lambda - \mathbf{A}_\lambda)\mathbf{m}^\perp dx_2 \right|^2 dx_1 \leq C\lambda^2, \quad (81)$$

that is,

$$\int_{-\ell}^0 \left| \int_{f_\lambda^+}^{\zeta+f_\lambda^-} (\mathbf{y}_{\lambda,2} - \mathbf{A}_\lambda \mathbf{m}^\perp) dx_2 \right|^2 dx_1 \leq C\lambda^2. \quad (82)$$

Integrating and using the boundary conditions (in the original variables, i.e., using the boundary conditions (54)), we have,

$$\begin{aligned} & \int_{-\ell}^0 \left| \int_{f_\lambda^+}^{\zeta+f_\lambda^-} (\mathbf{y}_{\lambda,2} - \mathbf{A}_\lambda \mathbf{m}^\perp) dx_2 \right|^2 dx_1 \\ &= \int_{-\ell}^0 |(f_\lambda^+ - f_\lambda^-)(\mathbf{B}_\lambda - \mathbf{I})\mathbf{m}_\lambda^\perp + (\zeta - (f_\lambda^+ - f_\lambda^-))(\mathbf{A}_\lambda - \mathbf{I})\mathbf{m}_\lambda^\perp|^2 dx_1 \leq C\lambda^2. \end{aligned} \quad (83)$$

Using once again the equations of the crystallographic theory, we have $(\mathbf{A}_\lambda - \mathbf{I})\mathbf{m}_\lambda^\perp = -\lambda(\mathbf{n} \cdot \mathbf{m}_\lambda^\perp)\mathbf{a}_\lambda$ and $(\mathbf{B}_\lambda - \mathbf{I})\mathbf{m}^\perp = (1 - \lambda)(\mathbf{n} \cdot \mathbf{m}_\lambda^\perp)\mathbf{a}_\lambda$. Thus, from (83) we get

$$\int_{-\ell}^0 |(f_\lambda^+ - f_\lambda^-) - \lambda\zeta|^2 dx_1 \leq C\lambda^2. \quad (84)$$

By the triangle inequality, $\|f_\lambda^+ - f_\lambda^-\|_{L^2} \leq \|f_\lambda^+ - f_\lambda^- - \lambda\zeta\|_{L^2} + \|\lambda\zeta\|_{L^2} \leq C\lambda$, we get

$$\int_{-\ell}^0 |f_\lambda^+ - f_\lambda^-|^2 dx_1 \leq C\lambda^2. \quad (85)$$

This can be combined with the L^∞ bound of the derivative by using the interpolation inequality ($f_\lambda = f_\lambda^+ - f_\lambda^-$),

$$\|f_\lambda\|_{L^\infty(\mathbb{R})} \leq C \|f_\lambda\|_{L^2(\mathbb{R})}^{\frac{2}{3}} \|f_\lambda'\|_{L^\infty(\mathbb{R})}^{\frac{1}{3}} \leq C c^{\frac{1}{3}} \lambda^{\frac{2}{3}}, \quad (86)$$

yielding the L^∞ bound,

$$|f_\lambda^+ - f_\lambda^-| \leq C\lambda^{2/3}. \quad (87)$$

8.4.4 Lower bound of the energy

We continue to assume that $\mathbf{y}_\lambda, f_\lambda^\pm$ is a low energy deformation, i.e., it has less energy than the naive test function (63), and we seek a lower bound for its energy. We change to the new variables $\mathbf{u}_\lambda(t_1, t_2)$ and define $\Omega_\ell = \{(t_1, t_2) : -\ell < t_1 < 0, 0 < t_2 < 1\}$. Let $\mathbf{v}_\lambda = \frac{1}{\lambda}\mathbf{u}_\lambda$. The scaled energy of $\mathbf{y}_\lambda, f_\lambda^\pm$ in the new variables is

$$\begin{aligned} & \frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda(\mathbf{v}_{\lambda,1} - g_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{m}_\lambda \\ & \quad + \lambda(\mathbf{a}_\lambda + h_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{n}) (\zeta - (f_\lambda^+ - f_\lambda^-)) dt_1 dt_2 \leq C, \end{aligned} \quad (88)$$

and we seek a good lower bound for the left hand side. First we note by the assumed lower bound $\tilde{\varphi}(\mathbf{F}) \geq C \text{dist}^2(\mathbf{F}, \text{SO}(3))$, the L^∞ bound on g_λ (cf., (58)), and the fact that the angle between the two vectors $\mathbf{A}_\lambda^{-T} \mathbf{m}_\lambda$ and $\mathbf{A}_\lambda^{-T} \mathbf{n}$ remains bounded away from zero in the limit $\lambda \rightarrow 0$ (as we have assumed), we have $\|\mathbf{v}_\lambda\|_{H^1} \leq C$ and therefore, for a subsequence (not relabeled), we have that $\mathbf{v}_\lambda \rightharpoonup \mathbf{v}$ in $H^1(\Omega_\ell)$.

We denote

$$\begin{aligned} \mathbf{G}_\lambda &= (\mathbf{v}_{\lambda,1} - g_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{m}_\lambda + (\mathbf{a}_\lambda + h_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{n}, \\ J_\lambda &= (\zeta - (f_\lambda^+ - f_\lambda^-)), \end{aligned} \quad (89)$$

and by (87) and the remarks just above, we have $J_\lambda \rightarrow \zeta$ uniformly and $\mathbf{G}_\lambda \rightarrow \mathbf{G}$ in $L^2((-\ell, 0) \times (0, 1))$. The bound (88) becomes

$$\frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) J_\lambda dt_1 dt_2 \leq C, \quad (90)$$

We now do a careful Taylor expansion of the integrand in the spirit of [12]. We introduce the “rare” set where \mathbf{G}_λ is large, i.e.,

$$E_\lambda = \{\mathbf{x} \in \Omega_\ell : |\mathbf{G}_\lambda(\mathbf{x})| \geq \lambda^{-\frac{1}{2}}\} \quad (91)$$

and we note that on $\Omega_\ell \setminus E_\lambda$, we have $\lambda|\mathbf{G}_\lambda| \leq \lambda^{1/2} \rightarrow 0$. Hence, using that $\tilde{\varphi} \in C^2$ and is minimized at \mathbf{I} , we have on $\Omega_\ell \setminus E_\lambda$,

$$\tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) = \frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2} (\lambda \mathbf{G}_\lambda, \lambda \mathbf{G}_\lambda) + \omega(|\lambda \mathbf{G}_\lambda|) |\lambda \mathbf{G}_\lambda|^2, \quad (92)$$

where $\omega(s) \rightarrow 0$ as $s \rightarrow 0$. Since we are only seeking a lower bound and $\tilde{\varphi} \geq 0$ we simply throw away the set E_λ , and write,

$$\begin{aligned} \frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) J_\lambda dt_1 dt_2 &\geq \frac{1}{\lambda^2} \int_{\Omega_\ell \setminus E_\lambda} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\lambda \mathbf{G}_\lambda, \lambda \mathbf{G}_\lambda) + \omega(|\lambda \mathbf{G}_\lambda|) |\lambda \mathbf{G}_\lambda|^2 \right] J_\lambda dt_1 dt_2 \\ &= \int_{\Omega_\ell \setminus E_\lambda} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}_\lambda, \mathbf{G}_\lambda) + \omega(|\mathbf{G}_\lambda|) |\mathbf{G}_\lambda|^2 \right] J_\lambda dt_1 dt_2 \end{aligned} \quad (93)$$

Since \mathbf{G}_λ is bounded uniformly in $L^2(\Omega_\ell)$ and $\omega(|\mathbf{G}_\lambda|) \rightarrow 0$ uniformly, then the error term can be neglected:

$$\begin{aligned} \liminf_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) J_\lambda dt_1 dt_2 &\geq \liminf_{\lambda \rightarrow 0} \int_{\Omega_\ell \setminus E_\lambda} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}_\lambda, \mathbf{G}_\lambda) \right] J_\lambda dt_1 dt_2 \\ &= \liminf_{\lambda \rightarrow 0} \int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\chi_{\Omega_\ell \setminus E_\lambda} \mathbf{G}_\lambda, \chi_{\Omega_\ell \setminus E_\lambda} \mathbf{G}_\lambda) \right] J_\lambda dt_1 dt_2. \end{aligned}$$

The quantity $\chi_{\Omega_\ell \setminus E_\lambda}$ converges boundedly almost everywhere to χ_{Ω_ℓ} and therefore $\chi_{\Omega_\ell \setminus E_\lambda} \mathbf{G}_\lambda \rightharpoonup \mathbf{G}$ (i.e., the same limit as previously) in $L^2(\Omega_\ell)$. By the lower bound assumed on $\tilde{\varphi}$, we have $\partial^2 \tilde{\varphi}(\mathbf{I})/\partial \mathbf{F}^2$ is a positive-definite quadratic form on symmetric matrices, that is, it is a positive-definite elasticity tensor. Thus, using its weak lower semicontinuity, we have the desired lower bound,

$$\liminf_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) J_\lambda dt_1 dt_2 \geq \int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}, \mathbf{G}) \right] \zeta dt_1 dt_2. \quad (94)$$

It remains to identify the specific form of the limit \mathbf{G} . By the definition (89) of \mathbf{G}_λ we have

$$(\mathbf{v}_{\lambda,1} - g_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{m}_\lambda + (\mathbf{a}_\lambda + h_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{n} \rightharpoonup \mathbf{G} \quad \text{in } L^2(\Omega_\ell). \quad (95)$$

Since $h_\lambda \rightarrow 1/\zeta$ uniformly, the only term whose limit is not completely clear is $g_\lambda \mathbf{v}_{\lambda,2}$. This term has a uniform bound in $L^2(\Omega_\ell)$, so we can find its limit by testing against a function $\psi \in C_0^\infty(\Omega_\ell)$ and integrating by parts:

$$\begin{aligned} \int_{\Omega_\ell} g_\lambda \mathbf{v}_{\lambda,2} \psi dt_1 dt_2 &= \int_{\Omega_\ell} \frac{t_2 f_\lambda^{-'} + (1-t_2) f_\lambda^{+'}}{\zeta - (f_\lambda^+ - f_\lambda^-)} \mathbf{v}_{\lambda,2} \psi dt_1 dt_2 \\ &= - \int_{\Omega_\ell} \frac{f_\lambda^{-'} - f_\lambda^{+'}}{\zeta - (f_\lambda^+ - f_\lambda^-)} \mathbf{v}_\lambda \psi dt_1 dt_2 - \int_{\Omega_\ell} g_\lambda \mathbf{v}_\lambda \psi, 2 dt_1 dt_2. \end{aligned} \quad (96)$$

We claim that both of the latter integrals tend to zero. This is clearly true of the first one, as $f_\lambda^{-'} - f_\lambda^{+'}$ is bounded in $L^\infty(\Omega_\ell)$ and therefore has a weak* limit; this limit must be zero since $|f_\lambda^- - f_\lambda^+| \rightarrow 0$. This combined with the strong convergence $\mathbf{v}_\lambda \rightarrow \mathbf{v}$ in $L^2(\Omega_\ell)$ establishes the assertion. For the second integral, we look at the boundary conditions (61),

$$\lambda \mathbf{v}_\lambda(t_1, 0) = f_\lambda^+(t_1)(1-\lambda) \mathbf{a}_\lambda, \quad \lambda \mathbf{v}_\lambda(t_1, 1) = f_\lambda^-(t_1)(1-\lambda) \mathbf{a}_\lambda, \quad -\ell < t_1 < 0. \quad (97)$$

Since by the trace theorem \mathbf{v}_λ is bounded in $L^2(\partial\Omega_\ell)$, then the uniform limits of f_λ^\pm , and therefore weak* limits of both $f_\lambda^{\pm'}$ individually must be zero. This shows that the second integral tends to zero. Summarizing, we have that

$$\mathbf{G} = \mathbf{v}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a}_0 + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n}. \quad (98)$$

Note that f^\pm play no role in the lower bound, which later turns out to be the Γ -limit. The boundary values of \mathbf{v} determine the values of f^\pm .

Also, since \mathbf{v}_λ is bounded in $L^2(\partial\Omega_\ell)$, it follows from the boundary condition $\mathbf{v}_\lambda(-\ell, t_2) = \zeta(1 - t_2 + \eta_\lambda/\lambda)(1 - \lambda)\mathbf{a}_\lambda$ that (passing, if necessary, to a subsequence),

$$\frac{\eta_\lambda}{\lambda} \rightarrow \hat{\eta}. \quad (99)$$

We conclude that there are several vestiges of the boundary conditions on \mathbf{v} and the presence of f^\pm . These are: (i) $\mathbf{v}(t_1, 0) \parallel \mathbf{a}_0$, $\mathbf{v}(t_1, 1) \parallel \mathbf{a}_0$, (ii) $(\mathbf{v}(t_1, 0) - \mathbf{v}(t_1, 1)) \cdot \mathbf{a}_0 \geq 0$, (iii) $\mathbf{v}(0, t_2) = 0$, and (iv) $\mathbf{v}(-\ell, t_2) = \zeta(1 - t_2 + \hat{\eta})\mathbf{a}_0$.

8.4.5 Upper bound

For the upper bound in the Γ -convergence argument, we have to show that, given $\mathbf{v} \in H^1(\Omega_\ell)$ satisfying the boundary conditions $\mathbf{v}(t_1, 0) \parallel \mathbf{a}_0$, $\mathbf{v}(t_1, 1) \parallel \mathbf{a}_0$, $(\mathbf{v}(t_1, 0) - \mathbf{v}(t_1, 1)) \cdot \mathbf{a}_0 \geq 0$, $\mathbf{v}(0, t_2) = 0$, and $\mathbf{v}(-\ell, t_2) = \zeta(1 - t_2 + \hat{\eta})\mathbf{a}_0$ there are $\mathbf{v}_\lambda \in H^1(\Omega_\ell)$, $f_\lambda^\pm \in W^{1,\infty}(-\ell, 0)$ (Lipschitz constant c), $f^+ \geq f^-$, and $\eta_\lambda \in \mathbb{R}$ satisfying the boundary conditions

$$\begin{aligned} \lambda \mathbf{v}_\lambda(t_1, 0) &= f_\lambda^+(t_1)(1 - \lambda)\mathbf{a}_\lambda, \quad \lambda \mathbf{v}_\lambda(t_1, 1) = f_\lambda^-(t_1)(1 - \lambda)\mathbf{a}_\lambda, \quad -\ell < t_1 < 0, \\ \mathbf{v}_\lambda(0, t_2) &= 0, \quad \mathbf{v}_\lambda(-\ell, t_2) = \zeta(1 + \eta_\lambda/\lambda - t_2)(1 - \lambda)\mathbf{a}_\lambda, \quad 0 < t_2 < 1, \end{aligned} \quad (100)$$

such that

$$\frac{1}{\lambda^2} \int_{\Omega_\ell} \tilde{\varphi}(\mathbf{I} + \lambda \mathbf{G}_\lambda) J_\lambda dt_1 dt_2 \rightarrow \int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}, \mathbf{G}) \right] \zeta dt_1 dt_2, \quad (101)$$

where $\mathbf{G}_\lambda, \mathbf{G}$ are calculated from $\mathbf{v}_\lambda, \mathbf{v}$ in the usual ways:

$$\begin{aligned} \mathbf{G}_\lambda &= (\mathbf{v}_{\lambda,1} - g_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{m}_\lambda + (\mathbf{a} + h_\lambda \mathbf{v}_{\lambda,2}) \otimes \mathbf{A}_\lambda^{-T} \mathbf{n}, \\ \mathbf{G} &= \mathbf{v}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a} + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n}. \end{aligned} \quad (102)$$

If $\mathbf{v} \in C^\infty(\Omega_\ell)$ and satisfies the boundary conditions $\mathbf{v}(t_1, 0) \parallel \mathbf{a}_0$, $\mathbf{v}(t_1, 1) \parallel \mathbf{a}_0$, and $(\mathbf{v}(t_1, 0) - \mathbf{v}(t_1, 1)) \cdot \mathbf{a}_0 \geq 0$, $\mathbf{v}(0, t_2) = 0$, and $\mathbf{v}(-\ell, t_2) = \zeta(1 - t_2 + \hat{\eta})\mathbf{a}_0$ then we simply take $\mathbf{v}_\lambda = \mathbf{L}_\lambda \mathbf{v}$, where $\mathbf{L}_\lambda \in \mathbb{R}^{3 \times 3}$ is a linear transformation satisfying $\mathbf{L}_\lambda \mathbf{a}_0 = \mathbf{a}_\lambda$ and $\mathbf{L}_\lambda \rightarrow \mathbf{I}$ as $\lambda \rightarrow 0$, and we take $\eta_\lambda = \hat{\eta}/\lambda$. Due to the differentiability of \mathbf{v} on $\partial\Omega_\ell$ we have for λ sufficiently small that f_λ^\pm defined by (100) is Lipschitz with Lipschitz constant c . Then the conclusion (101) follows by direct calculation using strong convergence and classical Taylor expansion.

If, however, we only know that $\mathbf{v} \in H^1(\Omega_\ell)$ and satisfies the boundary conditions, then there is a slight problem with satisfying the Lipschitz conditions on f_λ^\pm obtained from the (100). However, in this case we first approximate \mathbf{v} by $\mathbf{v}^{(\delta)} \in C^\infty(\Omega_\ell)$, $\mathbf{v}^{(\delta)} \rightarrow \mathbf{v} \in H^1(\Omega_\ell)$ and $\mathbf{v}^{(\delta)}$ satisfying the

same boundary conditions as \mathbf{v} . Using the quadratic growth of the integrand on the right and side of (101), we have that

$$\int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}^{(\delta)}, \mathbf{G}^{(\delta)}) \right] \zeta dt_1 dt_2 \rightarrow \int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}, \mathbf{G}) \right] \zeta dt_1 dt_2, \quad (103)$$

where $\mathbf{G}^{(\delta)} = \mathbf{v}^{(\delta)}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a}_0 + \frac{1}{\zeta} \mathbf{v}^{(\delta)}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n}$. For the left hand side of (101) we now choose $\mathbf{v}_\lambda = \mathbf{L}_\lambda \mathbf{v}^{(\delta)}$, $\lambda < \lambda_\delta$, with $\lambda_\delta > 0$ for $\delta > 0$ chosen small enough to respect the Lipschitz conditions. In particular, there are positive sequences $\lambda_k = \lambda_{\delta_k} \rightarrow 0$, $\delta_k \rightarrow 0$ such that $\mathbf{v}_{\lambda_k} = \lambda_k \mathbf{L}_{\lambda_k} \mathbf{v}^{(\delta_k)}$ satisfies the Lipschitz conditions. By Taylor expansion, strong convergence and (103) we have that (101) is satisfied for the sequence \mathbf{v}_{λ_k} .

8.5 Summary of form of the limiting energy of the transition layer

The lower bound found above in Subsection 8.4.4 shows that the limiting energy of a low energy sequence parameterized by λ is greater than or equal to

$$\int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{G}, \mathbf{G}) \right] \zeta dt_1 dt_2, \quad (104)$$

with \mathbf{G} replaced by $\mathbf{v}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a}_0 + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n}$. We show in Subsection 8.4.5 that given a function $\mathbf{v} \in H^1(\Omega_\ell)$ satisfying the limiting boundary conditions, its energy (104) can be approximated arbitrarily closely by the original expression for the energy evaluated on an appropriate sequence \mathbf{v}_λ satisfying the original boundary conditions. These two statements imply that the energy (104) is the Γ -limit of the original ($\lambda > 0$) energy under its boundary conditions.

In particular, it follows that the limit of the family of minimizers of the original energy, parameterized by λ , is a minimizer of (104). Also, the energy of this family, rescaled by dividing by λ^2 , converges to the energy of the Γ -limit.

The full statement of the Γ -limiting problem is:

$$\begin{aligned} \min_{\mathbf{v} \in \mathcal{A}} \int_{\Omega_\ell} \left[\frac{1}{2} \frac{\partial^2 \tilde{\varphi}(\mathbf{I})}{\partial \mathbf{F}^2}(\mathbf{v}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a}_0 + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n}, \right. \\ \left. \mathbf{v}_{,1} \otimes \mathbf{A}_0^{-T} \mathbf{m}_0 + (\mathbf{a}_0 + \frac{1}{\zeta} \mathbf{v}_{,2}) \otimes \mathbf{A}_0^{-T} \mathbf{n} \right] \zeta dt_1 dt_2, \end{aligned} \quad (105)$$

where

$$\begin{aligned} \mathcal{A} = \{(\mathbf{v}, \hat{\eta}) \in H^1(\Omega_\ell) \times \mathbb{R} : \mathbf{v}(t_1, 0) \parallel \mathbf{a}_0, \mathbf{v}(t_1, 1) \parallel \mathbf{a}_0, (\mathbf{v}(t_1, 0) - \mathbf{v}(t_1, 1)) \cdot \mathbf{a}_0 \geq 0, \\ \mathbf{v}(0, t_2) = 0, \mathbf{v}(-\ell, t_2) = \zeta(1 - t_2 + \hat{\eta})\mathbf{a}_0\}. \end{aligned} \quad (106)$$

Observe that $\hat{\eta}$, which describes the asymptotic lowering of the thin twin band, takes part in the minimization.

Acknowledgment. We wish to acknowledge valuable discussions with John Ball, Rémi Delville, Eckhard Quandt, Nick Schryvers, Doron Shilo and Thomas Waitz and the assistance of Sakthivel Kasinathan. This work was supported by the ARO-MURI W911NF-07-1-0410, the MULITIMAT RTN network MRTN-CT-2004-505226, AFOSR (GameChanger, GRT00008581/RF60012388) and DOE DE-FG02-05ER25706. The work also benefited from the support of NSF-NIRT DMS-0304326, ONR-MURI N000140610530, NIH Z196601 and the Humboldt Foundation.

References

- [1] J.M. Ball and C. Carstensen. Nonclassical austenite-martensite interfaces. *J. de Physique IV*, **7**(C5) (1997), 35-40.
- [2] J. M. Ball, C. Chu and R.D. James, Hysteresis during stress-induced variant rearrangement, *J. de Physique III*, Colloque C8, **5** (1995), 245–251
- [3] J. M. Ball and R. D. James, Fine phase mixtures as minimizers of energy. *Arch. Rational Mech. Anal.* **100** (1987), 13-52.
- [4] Giorgio Bertotti and Isaak Mayergoyz, The Science of Hysteresis. Academic Press (2006).
- [5] Kaushik Bhattacharya, Microstructure of Martensite. Oxford University Press (2003).
- [6] Z. Bo and D. Lagoudas, Thermomechanical modeling of polycrystalline SMAs under cyclic loading, parts I – IV. *Int. J. Eng. Sci.* **37** (1999), 1089-1140, 1141-1173, 1175-1203, 1205-1249.
- [7] J. S. Bowles and J. K. Mackenzie, The crystallography of martensitic transformation I and II. *Acta Metall.* **2** (1954) 129-137, 138-147.
- [8] J. Cui, Y. S. Chu, O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang and I. Takeuchi, Combinatorial search of thermoelastic shape memory alloys with extremely small hysteresis width. *Nature Materials* **5** (2006), 286-290.
- [9] E. De Giorgi and T. Franzoni, Su un tipo di convergenza variazionale. *Atti Accad. Naz. Lincei Rend. Sci. Fis. Natur.* **58** (1975), 842-850.
- [10] Rémi Delville, Richard D. James, Zhiyong Zhang and Dominique Schryvers, TEM investigation of twinning in ternary NiTiX alloys with special lattice parameters, Proc. ICOMAT-2008, accepted.
- [11] H.C. Donkersloot and J.H.N. van Vucht, Martensitic transformations in gold-titanium, palladium-titanium and platinum-titanium alloys near the equiatomic composition. *Journal of the Less-Common Metals* **20** (1970), 83–91.
- [12] Gero Friesecke, Richard D. James and Stefan Müller, A theorem on geometric rigidity and the derivation of nonlinear plate theory from three dimensional elasticity, *Comm. Pure and Appl. Math.* **55** (2002), 1461-1506.
- [13] K. Gall and H. J. Maier, Cyclic deformation mechanisms in precipitated NiTi shape memory alloys, *Acta. Materialia* **50** (2002), 4643–4657.
- [14] K. F. Hane and T. W. Shield, Microstructure in a cubic to orthorhombic transition. *J. Elasticity* **59** (2000), 267–318.
- [15] H. Horikawa, S. Ichinose, K. Morii, S. Miyazaki, and K. Otsuka, Orientation dependence of $\beta_1 \rightarrow \beta'_1$ stress-induced martensitic transformation in Cu-Al-Ni alloy. *Metallurgical Transactions* **19A** (1988), 915–923.

- [16] Xiangyang Huang, Graeme J. Ackland and Karin M. Rabe, Crystal structures and shape-memory behaviour of NiTi Nature Materials **2** (2003), 307-311.
- [17] R. D. James and Zhiyong Zhang, A way to search for multiferroic materials with “unlikely” combinations of physical properties, in *The Interplay of Magnetism and Structure in Functional Materials* (ed. L. Manosa, A. Planes, A. B. Saxena), Springer Series in Materials Science **79** (2005).
- [18] H. Kato, T. Ozu, S. Hashimoto and S. Miura, Cyclic stress-strain response of superelastic Cu-Al-Mn alloy single crystals, Materials Science and Engineering A **264** (1999), 245–253.
- [19] R. V. Kohn, R. D. James and T. W. Shield, Modeling the branched needle microstructures at the edge of a martensite laminate. *J. de Physique, Colloque C8*, vol. **5** (1995), 253-259.
- [20] R. V. Kohn and S. Müller, Branching of twins near an austenite-twinned-martensite interface. *Phil. Mag.* **66A** (1992), 697-715.
- [21] R. V. Kohn and S. Müller, Surface energy and microstructure in coherent phase transitions. *Comm. Pure Appl. Math.* **47** (1994), 405-435.
- [22] S. Kustov, S. Golyandin, K. Sapozhnikov, J. Pons, E. Cesari and J. Van Humbeeck, Effect of off-stoichiometry on the mobility of point-like defects and damping in binary CuAl martensites. *Acta Materialia* **54** (2006), 2075–2085.
- [23] P. H. Leo and T. W. Shield, Transient Heat Transfer Effects on the Pseudoelastic Behavior of Shape-Memory Wires, *Acta metallurgica* **41** (1993), 2477-2485.
- [24] C. LExcellent and G. Bourbon, Thermodynamical model of cyclic behaviour of TiNi and CuZnAl shape memory alloys under isothermal undulated tensile tests, *Mech. Mater.* **24** (1996), 59-73.
- [25] David S. Lieberman, Crystal geometry and mechanisms of phase transformations in crystalline solids, in Phase Transformations. American Society for Metals: Menlo Park, Ohio. (1968), 1-57.
- [26] T. Lim and D. McDowell, Cyclic thermomechanical behavior of a polycrystalline pseudoelastic shape memory alloy, *Journal of the Mechanics and Physics of Solids* **50** (2002), 651-676.
- [27] N.M. Matveeva, Yu.K. Kovneristyi, A.S. Savinov, V.P. Sivokha, and V.N. Khachin. Martensitic transformations in the TiPd-TiNi system. *Journal de Physique*, C4 (1982), 249–253.
- [28] S. Miyazaki and A. Ishida, Martensite transformation and shape memory behavior in sputter deposited TiNi-base thin films. *Materials Science & Engineering A*, 273-275 (1999), 106–133.
- [29] W.J. Moberly. Mechanical Twinning and Twinless Martensite in Ternary $\text{Ti}_{50}\text{Ni}_{(50,x)}\text{M}_x$ Intermetallics. Thesis, Stanford University (1991).
- [30] W.J. Moberly and K.N. Melton. Ni-Ti-Cu shape memory alloys, in Engineering Aspects of Shape Memory Alloys (ed. T.W. Duerig, K.N. Melton, D. Stöckel, and C.M. Wayman): Butterworth-Heinemann (1990), p. 46–57.

- [31] Z. Moumni, A. Van Herpen and P. Riberty, Fatigue analysis of shape memory alloys: energy approach. *Smart Mater. Struct.* **14** (2005), S287–S292.
- [32] R.C. O’Handley, D.I. Paul, M. Marioni, C.P. Henry, M. Richard, P.G. Tello and S.M. Allen, Micromagnetics and micromechanics of Ni-Mn-Ga actuation. *J. Phys. IV France* bf 112 (2003), 973-976.
- [33] K. Otsuka and C. M. Wayman (ed.), Shape Memory Materials, Cambridge University Press, 1998.
- [34] Francisco-José Pérez-Reche, Lev Truskinovsky, Giovanni Zanzotto, Training-induced criticality in martensites. *Phys Rev Lett.* **99** (2007), 075501.
- [35] P.L. Potapov, A.V. Shelyakov, A.A. Gulyaev, E.L. Svistunova, N.M. Matveeva, and D. Hodgson. Effect of Hf on the structure of Ni-Ti martensitic alloys. *Materials Letters* **32** (1997), 247–250.
- [36] Eckhard Quandt, Data on hysteresis and stress relaxation in films. Private communication (2007).
- [37] X. Ren, private communication (2007).
- [38] D. Schryvers, Microtwin sequences in thermoelastic $\text{Ni}_x\text{Al}_{1-x}$ martensite studied by high-resolution transmission electron microscopy. *Phil. Mag.* **A68** (1993), 1017-1032.
- [39] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, Hysteresis and hierarchies: dynamics of disorder-driven first-order phase transformations. *Phys. Rev. Lett.* **70** (1993), 3347–3350.
- [40] T. Shield, Orientation dependence of the pseudoelastic behavior of single crystals of Cu-Al-Ni in tension, *Journal of the Mechanics and Physics of Solids* **43** (1995), 869–895.
- [41] J. A. Shaw and S. Kyriakides, Thermomechanical aspects of NiTi. *Journal of the Mechanics and Physics of Solids* **43** (1995), 1243-1281.
- [42] D. Shilo, A. Mendelovich and V. Novák, Investigation of twin boundary thickness and energy in CuAlNi shape memory alloy. *Appl. Phys. Lett.* **90** (2007), 193113.
- [43] K. Tanaka, F. Nishimura, T. Hayashi, H. Tobushi and C. Lexcellent, Phenomenological analysis on subloops and cyclic behavior in shape memory alloys under mechanical and/or thermal loads, *Mech. Mater.* **19** (1995), 281-292.
- [44] P. Thamburaja and L. Anand, Polycrystalline shape-memory materials: effect of crystallographic texture. *Journal of the Mechanics and Physics of Solids* **49** (2001), 709–737.
- [45] T. Waitz, D. Spišák, J. Hafner and H. P. Karnthaler, Size-dependent martensitic transformation path causing atomic-scale twinning of nanocrystalline NiTi shape memory alloys. *Europhys. Lett.* **71** (2005), 98-103.

- [46] E. Vives, E. Obradó and A. Planes, Hysteresis and avalanches in the site-diluted Ising model: comparison with experimental results in CuAlMn alloys. *Physica B* **275** (2000), 45–49.
- [47] M. S. Wechsler, D. S. Lieberman and T. A. Read, On the theory of formation of martensite *Trans. AIME J. Metals* **197** (1953), 1503–1515.
- [48] W. Zaki and Z. Moumni. A 3D model of the cyclic thermomechanical behavior of shape memory alloys. *Journal of the Mechanics and Physics of Solids* **55**, 2427–2454.
- [49] W. Zhang and K. Bhattacharya, A computational model of ferroelectric domains. Part II: Grain boundaries and defect pinning. *Acta Materialia* **53** (2005), 199–209.
- [50] Zhiyong Zhang, Special Lattice Parameters and The Design of Low Hysteresis Materials, Thesis, University of Minnesota (2007).
- [51] Zhiyong Zhang, Private communication on the effect of stress on lattice parameters and hysteresis (2008).